REPORT RESUMES

ED 010 350 48
LANGUAGES OF THE WORLD--SINO-TIBETAN FASCICLE ONE.
BY- VOEGELIN, C. F. VOEGELIN, FLORENCE M.
INDIANA UNIV., BLOOMINGTON
REPORT NUMBER NDEA-VI-63-3 PUB DATE MAR 64
CONTRACT OEC-SAE-9488
EDRS PRICE MF-$0.18 HC-$4.60 115P. ANTHROPOLOGICAL
LINGUISTICS, 6(3)/1-109, MARCH 1958

DESCRIPTORS- *SINO TIBETAN LANGUAGES, *LANGUAGES, CHINESE, BLOOMINGTON, INDIANA, ARCHIVES OF LANGUAGES OF THE WORLD

THE CHINESE LANGUAGE AND ITS DIALECTS ARE DESCRIBED. (THIS REPORT IS PART OF A SERIES, ED 010 350 TO ED 010 367.) (JK)

ED010350

63-3

U. S. DEPARTMENT OF HEALTH, EDUCATION AND WELFARE
Office of Education
This document has been reproduced exactly as received from the person or organization originating it. Points of view or opinions stated do not necessarily represent official Office of Education position or policy.

Anthropological Linguistics

Volume 6 Number 3

March 1964

Title as is

LANGUAGES OF THE WORLD:
SINO-TIBETAN FASCICLE ONE

A Publication of the
ARCHIVES OF LANGUAGES OF THE WORLD
Anthropology Department
Indiana University

ANTHROPOLOGICAL LINGUISTICS is designed primarily, but not exclusively, for the immediate publication of data-oriented papers for which attestation is available in the form of tape recordings on deposit in the Archives of Languages of the World. This does not imply that contributors will be restricted to scholars working in the Archives at Indiana University; in fact, one motivation for the publication of ANTHROPOLOGICAL LINGUISTICS is to increase the usefulness of the Archives to scholars elsewhere by making publishable data and conclusions and their tape recorded attestation more widely available. (Recorded attestation of papers from scholars elsewhere will be copied by the Archives and the original recordings returned to the collector; others may then work with the tape copy either in the Archives or elsewhere by having a copy sent to them.) In addition to heavily exemplified papers in the form of preliminary or final statements on restricted problems in phonemics, morphophonemics, morphemics, syntax and comparative grammar, ANTHROPOLOGICAL LINGUISTICS will include lexical lists and analyzed texts (especially in the otherwise hard-to-publish range of 20 to 100 pages) and theoretical and methodological papers (especially in the form of papers from symposia).

Each volume of ANTHROPOLOGICAL LINGUISTICS consists of nine numbers to be issued during the months of January, February, March, April, May, June, October, November and December. Subscriptions ($3.50 a year) and papers for publication should be sent to the editor, Dr. Florence M. Voegelin, Anthropology Department, Indiana University, Bloomington, Indiana.

Entered as second class matter at the postoffice
at Bloomington, Indiana

LANGUAGES OF THE WORLD:
SINO-TIBETAN FASCICLE ONE

C. F. and F. M. Voegelin

Indiana University

0.1. Authorship and sponsorship

0.2. Scope of Sino-Tibetan

0.3. Previous Sino-Tibetan classifications

1.0. Han Chinese as an ethnolinguistic monolith

1.1. Character writing

1.2. Han Chinese in conversation and in Wenyen

1.3. Common national language = 'acquired Mandarin'

1.4. Sentence Profiles of 'acquired Mandarin'

1.5. Diversity among Chinese languages and dialects

The research reported herein was performed pursuant to a contract with the United States Office of Education, Department of Health, Education, and Welfare.

0.1. This is the first Fascicle of the Languages of the World File — an NDEA Office of Education project begun in Washington under the directorship of William R. Parker and Kenneth Mildenberger. The first place that the Language File was developed was in the Office of Education, Washington, D. C., in response to a practical need. Various administrative and legislative offices in Washington found it difficult to obtain reliable information on the languages of particular countries, as the number of different languages and the number of their speakers (language census), and whether two different language names represented dialect differences of one language, or a language barrier to each other, and so on. Consultants were invited to come to Washington to supply such information within the areas of their competence; and the information they gave was placed in a manuscript Language File.

This Language File was then moved from the Office of Education to George Washington University in Washington, D. C. where it was developed along the same lines, and in addition employed a small staff which excerpted external information from published sources, especially information on language census. This permitted the compilation of a list of languages having over a million speakers, subsequently published with a preface by the principal investigators of the present Fascicle who pointed out that a census giving the number of speakers ascribed to language names in the published sources could be misleading (Languages Now Spoken by Over a Million Speakers, AL 3.8.13-22, 1961). For example, in the published list the maximum number of speakers for any one language in South Asia is ascribed to Bengali (75 million), since Hindi is listed separately from Urdu: but the combined number of speakers of the one language, Hindi-Urdu (135 million), far exceeds the number of speakers of Bengali. A reliable language census is possible only when external information about the languages is combined with internal information on the languages themselves.

Finally, the Language File was moved from Washington, D. C. to Indiana University, which continued its development in the following fashion. First the aid of Corresponding Contributors was elicited — specialists were asked to supply bibliography relevant to Sino-Tibetan languages, for example, as well as to bibliographic sources for such information. Second, a staff of about a dozen advanced graduate students in linguistics excerpted the published sources and placed the full citations and reduced abstracts of what they excerpted under the appropriate rubrics in the Language File. Third, the excerptions were organized and summarized by the principal investigations for critical appraisal and revision by the consultants, either at Indiana University or at the university of the consultant.

This report is Englished by the principal investigators who wish to make it clear that the report is, in intent, equally authored by all the consultants. In our original plan, a final typescript of this report was to be sent to each consultant for his addenda and corrigenda. Some additional material was obtained in this way, and integrated in the report as it now stands. However, changes were also made without approval of all the consultants listed below and in all cases the integration of corrections were made without subsequent perusal and approval of the consultants who contributed the correction. It would seem improper to imply formal responsibility of joint authors to consultants who did not have time to examine the final version of this report as it now stands; and though we were willing to give consultants all the time that was needed, the terms of our contract called for completion of our contract on this and following fascicles by a date which could not be further extended. Hence, we list the consultants of FASCICLE ONE here instead of under the title of FASCICLE ONE, as was our original intention. The readers of this report as well as the principal investigators (C. F. and F. M. Voegelin) have the enormous benefit of co-authorship by the following consultants: Y. R. Chao, of the University of California, Berkeley; Nickolas Bodman, of Cornell University, who reviewed material presented here but was mainly concerned with materials to appear in subsequent fascicles, as was Denzel Carr of the East-West Center at the University of Hawaii, and the University of California, and as was Robert Morse who worked at Indiana University but is now re-engaged in field work in Burma; Henrietta Chen, also formerly at Indiana University but now at Cornell University, is the source of what we call 'acquired Mandarin', and also the informant for the Thai sentences which appear in a following fascicle. As matters stand, this is a very preliminary rather than a polished publication on the state and scope of Sino-Tibetan. The principal investigators request specialized scholars to supply them with corrigenda and addenda which may be published, as such, or incorporated in a future revised version of the present report.

Our development of the Language File included excerpting and critical consideration not only of external information, as language census data, but also of internal information on the languages themselves — partly in order to control and thus add to the reliability of external data, partly to make the fascicles relevant to the broader interests of administrators, legislators, and scholars who are not linguists, but who are practically or theoretically concerned with communicational complexities existing in the world today.

The internal information about the languages themselves is given in non-technical observations after illustrative sentences — observations on whether the structure illustrated is productive (generative), i.e. widely applicable to other sentences with different selections of topics, comments, or phrases, for example; observations on how the structure illustrated is like or unlike other areas of the same grammar or similar grammars, as Thai in reference to Chinese; and observations about different functions fulfilled by a particular structure illustrated. Each illustrative sentence in the source language is flanked by English translations — preceded by a free translation and followed by a literal translation in which the glosses in English follow the order of 'words' in Chinese, for example.

The set of all sentences for each language or dialect group is then followed by a sketch of its sound system. Here we give little more than the value of the letters (phonemes) in which the preceding illustrative sentences were written.

But for the higher grammatical levels, we present samples from the language — the illustrative sentences — which the observant reader may understand directly and immediately. Our remarks appended to these illustrative sentences are offered in the spirit of observations for the observant. And if the observant reader, guided by the illustrative sentences, anticipates our appended observation, so much the better. In short, this presentation is intended to give the reader internal or grammatical information, about Chinese, for example, derived from his own observation of a sample of Chinese.

All this might be called the presentation of grammar without linguistics — but not without benefit from linguistics. We are all indebted to structural linguistics for justifying the particular structure of a particular language. It is the justification that is hard to follow for the non-linguist — not the habits of the speakers of a particular language. Grammars are justified statements about those habits. After all, every Chinese child learns to follow the habits of his predecessors in a particular Chinese dialect, without linguistic training. But, of course, the child is only an exemplar of such habits, and does need linguistic training before he can make valid statements about them. When the ethnolinguistic or sociocultural factor is included, then language as the product of speakers' habits appears in an expanded frame of reference — the functional frame that Ferguson has called the language situation in a particular country.

0.2. However, if we were to begin with the language situation in China and go on to the language situation in other countries - in Tibet, and South Asia and much of Southeast Asia in which Sino-Tibetan languages are spoken - we should encounter so many unrelated languages that the picture of Sino-Tibetan languages would become blurred, and the list of Sino-Tibetan languages given in indexes to the present fascicle would have to be taken out of a welter of all the language names that are encountered in the countries in which Sino-Tibetan languages are also spoken. We can still refer to these other languages - languages not related to Sino-Tibetan - but in terms of their language family labels (as Dravidian or Indo-European or Malayo-Polynesian) rather than in terms of their individual language names. We are thereby enabled to focus on Sino-Tibetan languages, and list them separately as individual languages, though in the context of the countries in which they are spoken. In successive fascicles, the other languages of the countries will be listed separately, with the cumulative result of obtaining a list of all languages for all countries.

Here we are sufficiently concerned with Sino-Tibetan languages. How many are there? Limiting ourselves to language names about which some linguistic data is published or is in accessible manuscript form, we obtain a list of several hundred language names. How many people speak Sino-Tibetan languages and where are they located? The speakers of these languages are very unevenly distributed. Most are found in mainland China, whose population has now reached 750 million — including non-Chinese speakers — according to information vouchsafed to President Kennedy recently (December 17, 1962, televised press conference). In Burma there are 16 million speakers of Sino-Tibetan languages, and in other parts of Southeast Asia, as Thailand and Laos, such speakers are also counted by the millions. In addition, people in northern India (Nepal, Sikkim, Bhutan, and northern Burma), speak numerous Sino-Tibetan languages which average about 180 thousand speakers for each language. In terms of density of Sino-Tibetan speakers, mainland China is enormously dense, while Tibet and South Asia are more sparsely represented. But in Southeast Asia, the number of Sino-Tibetan speakers is again relatively dense, especially when the overseas Chinese are counted with speakers of national languages, as Thai in Thailand. The overseas Chinese in Singapore outnumber the native Malayo-Polynesian speakers in Singapore, for example.

The family tree model is more appropriate for the different language families in the Sino-Tibetan phylum than for the phylum itself. Each language family treated here is taken as a constituent of the Sino-Tibetan phylum. This and following fascicles list languages in the following nine language families: 1. Han Chinese; 2. Miao-Yao; 3. Kam-Thai; 4. Burmese-Lolo; 5. Karen; 6. Bodo-Naga$_1$-Kachin; 7. Naga$_2$-Chin; 8. Gyarung-Mishmi; 9. Tibetan. Each of these nine language families are reconstructed (or apparently reconstructible) for nine different parent languages. After these are actually reconstructed, their relationship to each other can be appreciated in Sino-Tibet phylum linguistics, rather than controverted. We avoid the controversy by giving exclusive attention to the languages spoken today in each of the constituent language families. That is to say, we wish to avoid becoming entangled in the controversy over competing classifications which we outline in the following section (0.3.).

0.3. There are over 30 million speakers of languages of the Kam-Thai family, of whom over 20 million speak languages belonging to the Southwestern branch — Thai, Lao, Shan, Thai Noir, Paiyi; in addition Maspero (1912, 1952) postulated Vietnamese as a member of this family.

Disturbing questions now obtrude. Does Kam-Thai, for example, belong to the Sino-Tibetan family, if Sino-Tibetan is a language family ? Or does Kam-Thai belong with Malayo-Polynesian languages ? There is an interesting history in this controversy over classification. Wilhelm Schmidt's classification includes Kam-Thai in Sino-Tibetan and in addition postulates Austric as ancestral to Austronesian (Malayo-Polynesian) and to an Austroasiatic phylum (including Munda and Mon Khmer) and then postulates a relationship between Sino-Tibetan and Austric—thereby making Malayo-Polynesian languages in the Pacific related to South, Southeast and East Asian languages. George Grierson's classification starts with Sino-Tibetan as the earliest ancestor, and has the daughter languages divided into Thai-Chinese languages and Tibeto-Burman languages. P. K. Benedict begins with bifurcating branches as does Grierson, but has the daughter languages divided into Tibeto-Karen languages and Chinese languages - rather than Thai-Chinese. This is because in Benedict's view, the Thai languages are daughter languages of another parent language, also called Austric, which also bears an ancestral relation to Mon-Khmer, Vietnamese, and Malayo-Polynesian languages (as well as to his 'Kadai', which we include in Kam-Thai). This differs from the first classification noted above chiefly in that two different parent languages are postulated by Benedict (Sino-Tibetan and Austric) while one parent was previously postulated — ancestral to Mon-Khmer and to Malayo-Polynesian languages as well as to Sino-Tibetan, with the latter parental to Thai-Chinese. Most subsequent summaries of languages of the world follow either the more inclusive classification (e.g., L. H. Gray), or the less inclusive Grierson classification (e.g. E. Kieckers); one recent summary (Greenberg's) seems to have followed the least inclusive Benedict classification, which does not even include Thai among the Sino-Tibetan languages. The fundamental issue here does not hinge on the present-day data (since everyone can see the plethora of similarities between Chinese and Thai), but on the interpretation of these similarities in grammar and lexicon. If the similarities are thought to reflect a reconstructible Proto Thai-Chinese, then Sino-Tibetan is still ancestral to Thai-Chinese but more distantly so. But if the similarities between Chinese and Thai are thought to be a consequence of diffusion between the two, then the problem is shifted to areal linguistics, which treats questions of borrowing between languages that are not necessarily related.

This might end in controversy if our alternatives were to accept or reject Sino-Tibetan as the parent language of the languages variously classified under it. Controversy is avoidable, however, if Sino-Tibetan is postulated as the non-reconstructed phylum ancestor of a number of language families, each with its own reconstructed (or obviously reconstructible) parent language. No one can quarrel with setting up Sino-Tibetan as a cover term for languages that bear some similarity to each other, beyond random similarity. If Thai does not belong in this phylum (as Benedict says it does not), it belongs in another phylum in the Pacific, with a different phylum ancestor; then it might be said (as Schmidt has said or rather implied) that a macro-phylum ancestor (Austric) does ultimately relate both the Sino-Tibetan phylum and the Thai languages. For further details of the differences between the Schmidt, Grierson, and Benedict classification and Shafer's summary, see the appended classification charts. The differences among these classification charts suggests that fruitless controversy would follow if we were to regard Sino-Tibetan as one language family rather than as a linguistic phylum, which is ancestral to many language families. We look for a reconstructed or reconstructible parent language for each language family under the phylum. These more modest language families then account for the closer relationship between groups of languages shown nearer the bottom of each chart in a given family — nearer the present day differentiation — and the remoter relationship between language families, of which there are nine altogether, is postponed as a problem of Sino-Tibetan phylum linguistics.

This is essentially the direction in which both T.H. Tung and Y. R. Chao move in their treatment of Sino-Tibetan languages spoken within China and in Southeast Asia. They set up what amounts to a Chinese language family, a Miao-Yao language family, and a Kam-Thai language family; but they lump the remaining languages of the Sino-Tibetan phylum into a non-coordinate Tibeto-Burman branch of the main phylum. The latter is susceptible to language family grouping, just as Tung and Chao have set up Chinese, Miao-Yao and Kam-Thai in East and Southeast Asia; and credit belongs to Robert Shafer for the first glimmering of this in his attempt to break up the non-coordinate Tibeto-Burman branch into coordinate language families.

The breakthrough in finding order among the Sino-Tibetan languages comes with the increased recognition that we are actually dealing with more or less distantly related language families, each of which has to be described before agreement can be reached in Sino-Tibetan phylum linguistics. The family tree model for the Sino-Tibetan phylum given in all previous classifications postulates more than is known and hence ends in a list of non-agreeing charts, such as those given below — or even includes an American Indian language family (Athapascan). What is known is that there is more similarity among Sino-Tibetan language families than among any randomly selected group of nine or ten language families in any part of the world. What is not yet known is how much of the similarity among Sino-Tibetan language families is reconstructible, and how much is a consequence of diffusion, possible because of the coexistence of Sino-Tibetan language families in one vast area extending from north India and from Southeast Asia through China into Manchuria.

A set of charts, and also further discussion, of the classification of Sino-Tibetan languages, according to Wilhelm Schmidt, according to George

Grierson, according to P. K. Benedict, according to the Chinese scholars cited and others, and according to Robert Shafer are given in subsequent fascicles.

1.0. Of all large countries, mainland China comes closer than any other to being monolithic or megalithic in the habits and appearance and numbers of its speakers. It is easier to understate than to overstate this aspect of the languages and populations of mainland China. To obtain a relative view, compare mainland China and India.

In population China is first in world rank; India is second. There are, roughly speaking, two billion people in all of Asia. Almost half of this vast population is Han Chinese. More than one quarter of the two billion people of Asia (some 600 million) live in South Asia — that is, in India, Pakistan, Ceylon, Nepal, Sikkim and Bhutan; and some geographers also include Afghanistan in South Asia.

Languages in one family of South Asia are related to English (Indo-European, hereinafter IE), but these languages, intrusive from Europe and West Asia in prehistoric times, have been influenced to sound in some respects like languages which were in India before the arrival of IE speakers — at least those of the Dravidian and Munda families. Still other language families, intrusive from Tibet and China, belong to the Sino-Tibetan phylum. And the total number of different or separate languages — each constituting a language barrier to the other (whether related in the same family or not) — is well over a hundred for South Asia. However, in mainland China, the vast majority of the 750 million inhabitants speak what we term Han Chinese. Some 95 per cent of the 750 million inhabitants speak one or another of the Han Chinese languages, and 70 per cent of the Han Chinese speak Mandarin;

the remaining 30 per cent speak Wu, Min, Hakka, Cantonese, and Hsiang. There are also eleven million Han Chinese speakers in the official Republic of China of the United Nations (=Nationalist China=Taiwan=Formosa)— a refugee government on part of its own territory. And additional millions of overseas Chinese speak Han Chinese— mostly in Southeast Asia.

Within China itself the non-Han speakers are also regarded as being non-Chinese. Some speak an unrelated foreign language, as English; many speak various tribal languages belonging to other (but only to two other) language families in what is sometimes called 'China proper' - Miao-Yao and Kam-Thai (cp: following fascicles). Both of these, like Han Chinese, are language family constituents of the Sino-Tibetan phylum.

Taiwan's population consists of a majority of speakers of Southern Min, perhaps half of whom can also use Mandarin, plus a minority of native speakers of Mandarin and now only a very small sprinkling of speakers of the non-Han tribal languages which belong to the Austronesian language family, which is not a constituent of the Sino-Tibetan phylum. The island of Hainan— under Communist control, though not part of mainland China— probably has fewer speakers of Han Chinese (specifically, Southern Min) than of Li which belongs to the Kam-Thai family (also a constituent of the Sino-Tibetan phylum), besides possibly speakers of Austronesian languages.

The notion of 'China proper'— in general area, extending south of the Great Wall— is convenient for distinguishing the part of China in which

Han Chinese languages are predominant from the rest of the country which may be called China's frontier provinces. The area covered by both 'China proper' and the frontier provinces is much greater than that of the United States (though 'China proper' alone is somewhat smaller). Less than a fourth of the inhabitants who speak a language which is Sino-Tibetan—but not Han Chinese—reside in 'China proper'; three-fourths of them reside in one or another of China's frontier provinces, and in some of these there also reside peoples who speak Altaic languages which are neither Han (in language family affiliation) nor Sino-Tibetan (in phylum affiliation).

Thus, there are 5 million Uighurs in Sinkiang province, speaking Altaic languages known as Kazak (I-li) and Khalkha (K'e-tzu-le-su), beside the Mohammedan Hui (Ch'ang-chi). There are more Mongol speakers in China's province of Inner Mongolia than there are in the Mongolian Republic. Most inhabitants of Tsinghai province and some in Ninghsia province, as well as in Manchuria, speak Altaic languages. The remaining frontier provinces include many languages which are non-Han Chinese but still Sino-Tibetan, as the three Tibetan languages in China's province of Tibet, with still other Sino-Tibetan languages represented in most of Yunnan province, most of Kwangsi province, and in parts of other provinces (Szechuan, Kweichow, and western Hunan). Nor are the Tibetan languages confined to the Tibetan province; they are also spoken in adjacent Yunnan where a total of 6 million people speak Sino-Tibetan languages that are not Han Chinese.

The population of all of South Asia (600 million) approximates that of Han Chinese speakers alone - variously estimated as 94 or 95 per cent of the total population, including speakers of all languages, both in 'China proper' and in the frontier provinces. Language census data in recent years show steady increase of 15 million or about 2% annually for the total population in China. Thus, the total figures published in 1948 give 463 million, those in 1953 give almost 583 million, and those in 1962 give 699 million, with an even higher total (750 million) given in President Kennedy's news release. If we estimate the Han Chinese speakers to be 95 per cent of these totals, then their numbers for 1948 are 440 million; those for 1953 are 554 million; those for 1962 are 664 million, or even more (710 million). However, when language census figures are itemized for particular Han Chinese languages or dialects, the sum for all Han Chinese speakers falls below rather than above 600 million (but some itemizations include more dialect-language names than others). We cite an itemization from a Chinese source (Lou Cheng-Pei and Lü Shu-Xiang, Yuyan Yanjiu, p. 4, 1956) which, though closely similar to a Russian source (G. P. Serdjučenko for 1959), does not include the dialect-language name Kan to which the Russian source ascribes 13 million speakers. The 1956 itemization speaks of dialects rather than languages; of the Hakka dialect (with 20 million speakers), of the two Min dialects (with 22 million speakers), of Hunan, i.e., Hsiang (with 26 million speakers), of Cantonese (with 27 million speakers), of Wu (with 46 million), and of Mandarin (387 million, including Northern, Southern, and Southwestern Mandarin).

The monolithic language situation, then, is surely the result of the numerical preponderance of Han Chinese speakers in China, but only partly so. Several other factors contribute to the same monolithic result, despite one factor which would seem to have the opposite effect. That one factor is that there is not one Chinese language; rather there are six or seven different Chinese languages. In this western sense, Han Chinese is a language family. But why do non-western Chinese scholars generally — and for good reasons, though not for the reasons adopted in western dialectology — speak of dialects rather than languages? In the Chinese view (as well as in the view of many western Sinologists) Hakka, Min, Hsiang, Cantonese, Wu, and Mandarin are dialects of a single Chinese language.

These Han Chinese dialects or languages are unified enthnolinguistically if not linguistically. Whether viewed as dialects of one language, or as separate languages in one language family, no one can quarrel with calling the whole 'Han Chinese'. Han Chinese is about as unified when taken as a cultural construct, as when taken as a linguistic construct.

There are cultural differences in mainland China, such as that based on rice subsistence in the South, and that based on wheat subsistence in the North; representatives of the latter are distinguished as 'noodle eaters' by Chinese students who come to America from South China. However, both North and South Chinese are equally Han. Also unifying is the stereotype of the Middle Kingdom Man; but this stereotype does not coincide with the referent of Han Chinese when it includes some non-Han Miao-Yao.

There are linguistic differences of a complete language barrier kind between Mandarin and Cantonese, for example, leading to the western view of separate languages — but all belonging to the Han Chinese language family. We do not treat the linguistic diversity of earlier times, but note in passing that linguistic diversity is not confined to modern China.

It is only in the ethnolinguistic view that the monolithic construct of Han Chinese emerges. This is the view that is given below. From a superficial point of view, it would appear that Han Chinese as an ethnolinguistic monolith has played into the hands of the Communists who succeeded in uniting mainland China, while in South Asia the splintering of languages and cultures has stood as a bulwark against unification. But the fact is that the Communists are playing down the unification influence of Han Chinese; what we call Han Chinese is called by them Hàn Yǔ, to distinguish Han Chinese from Miao-Yao, etc. The Communists' implication is that Han Chinese is only one of several tribes and languages in the Middle Country; the others, though minority tribes and languages of the Middle Country, are then regarded as being ethnically Chinese.

1.1. Writing in Chinese characters began at about the time that the Sumerian language became extinct — i.e. about 2000 B.C. Though Chinese is not the oldest known kind of writing, it is the oldest writing of its kind that has not become extinct. Other writing systems that were known in the Most Ancient East (from Sumeria to Egypt and Minoa) had to be deciphered after millenia of disuse; Maya glyphs (characters) in Mesoamerica are now being deciphered after centuries of disuse. In their continued use of the alphabet-included logographic system, the Chinese alone have kept alive a type of writing which no other culture was able to keep alive. All writing of this type uses some symbols (characters) to specify words or smaller parts of words (morphemes); and that is all that is meant by 'logographic'. Chinese characters are 'logographic' rather than 'ideographic'; they are word-symbols (or more exactly, morpheme-symbols) rather than idea-symbols. And they are certainly no more pictographic than English words are sound——imitative — in both cases, a few are; most are not.

Chinese newspapers keep several thousand characters on hand. The Chinese telegraphic code book is limited to 10 thousand characters. (This is not quite as arbitrary a closed corpus as it has sometimes been said to be. The telegraphic code provides 10,000 four-digit groups — 0000 to 9990 — for over nine thousand characters, with a few unused or blank spaces; and in addition there are available some graphic variants and obsolete characters.) And the fullest unabridged dictionary, begun in the 18th century, gives some 50 thousand characters. Attempts at simplifying character writing were made by early missionaries and by recent Communists. During the 20th century, Chinese scholars have also used adaptations of our self-sufficient alphabet with its low number of letters, besides continuing to use the thousands of characters required for writing in the alphabet-included logographic system. There are, accordingly, two co-existent but not competitive writing systems in modern China: the traditional characters used by all literate Chinese, and variants of the Latin alphabet as by specialized scholars. The Gwoyen Romatzyh (the National Romanization) and the new system used in Mainland China were intended to co-exist for a while with and later to supplant the Chinese characters as the national script, but have as yet not widely succeeded. The RussianCyrillic alphabet is used for writing Chinese by Russian and other Soviet scholars in much the same way as romanizations are used in the West.

In alphabetic type, however, the Russian Cyrillic alphabet is the same as the Latin alphabet. The latter was stabilized in its full inventory of 26 letters less than a thousand years ago, though the history of the Greco-Roman-Greco-Russian alphabetic type goes back to the time when the Greeks borrowed letters from the Phoenicians who were using a different type of alphabet. In Greek and later usage, letters specifying consonants appear beside letters specifying vowels. The use of symbols to specify sounds is all that is meant by alphabetic writing.

It is in this sense that the Chinese system of writing is said to be an alphabet-included logographic system. The alphabetic component of some characters gives a phonetic hint (sound specification) while the logographic component of the same character gives the lexical or morpheme specification. What is surprising to westerners is that the vast majority of Chinese characters do give some phonetic information.

This is so contrary to the western bias about Chinese writing that we invoke a fantasy to correct the bias — that the enormous difficulty in learning to read thousands and thousands of characters derives from the erroneous assumption that the characters are mostly without any phonetic hint as to their pronunciation. Suppose, now, that the Communist hegemony in mainland China had not merely threatened, but had persisted in its threat to replace the traditional alphabet-included logographic system with a self-sufficient alphabet which exists beside the former but does not now replace it. Suppose, then, that the Chinese characters had been entirely abolished in the second half of this century. Suppose further that a twenty-first century committee of Chinese linguists were charged with deciphering the alphabet-included logographic writing in one library of untranslated text manuscripts; how would they go about it? From the point of view of statistical preponderance, the characters which include sound specification would be the most important set in the decipherment of an interrupted tradition in writing. If the decipherers were to neglect the sound-specification components in such characters, they would have no more chance of success than Sir Arthur Evans had in deciphering Minoan writing; it was only when Michael Ventris took the phonetic part of Minoan writing as his point of departure that he was able to show that Minoan writing consisted of a juxtaposition of a self-sufficient alphabet (called Mycenean Linear B), and logograms (pictographic and quantifying logograms).

But Chinese character writing is not a system which juxtaposes a string of self-sufficient alphabetic symbols (specifying sounds) to a string of logographic symbols in Minoan-like equations. Rather, it is a system which integrates sound symbols and lexical morpheme symbols. The overwhelming majority of characters in the liu shu classes or categories of the Chinese lexicographers include sound specification. Two clear exceptions to the rule for the majority are the pictographic hsiang hsing class of characters, and the easy-to-remember 'simple ideograph' or chih shih class (e.g. one line marks one, two lines two, etc.); but both of these classes are numerically negligible. Slightly more numerous is the 'compound ideograph' class of characters that more nearly fits the western image of subtle Chinese thinking - e.g. the single character for military (combining constituents for caught in the middle + to stop + arms). Most numerous are classes of characters with alphabet-included constituents - i.e. with components that specify sounds. (But some of the latter do double duty simultaneously: some components specify both sound and morpheme.)

Why then is Chinese character writing not taken as another type of alphabetic writing for the most part? Would it not simplify matters to divide Chinese characters into two separate parts - one part made up of logographic classes (as *hsiang hsing* and *chih shih*) versus all the rest (those including sound specification in the characters)? No, because the alphabet-included part of a Chinese character is always integrated in a single composite character with the logographic part; the alphabet-included part specifies sounds, but not self-sufficiently so.

What are called simply 'alphabets' are really self-sufficient alphabets. Such alphabets are sometimes used with juxtaposed Chinese-like logograms, as the Arabic number 5, or ♀ for *female*. And 5 and ♀ are no doubt logograms because they specify words in any language that uses them, irrespective of the way the words sound: 5 = *five* in English but *beç* in Turkish. If we were to invert the question asked above in respect to Chinese writing (why not separate the alphabet-included part from the logographic part?), and address the question to alphabetic writing of any European language, as English, the question would be not unreasonable. Why not take English writing as a simpler but still Chinese-like type of alphabet-included logographic system, since the alphabetic inventory is limited to 26 letters (not counting calligraphic variants, as italic, boldface, gothic, etc.), and since the inventory of logograms is greater then 26 symbols if scientific symbols, such as ♀, are added to the more widely used Arabic numbers? There is one fact that precludes this inversion. Because the alphabetic part of English writing - like Mycenean Greek - is self-sufficient, it can dispense entirely with the logographic part (but not vice versa).

Hence English writing as a whole is a particular instance of a logograph-included (includable) alphabetic type.

Is there any hope of developing a more Chinese-like logographic system of writing for Indo-European languages? This would liberate the writer and reader from mastering the different IE languages, and still enable him to communicate in writing across language barriers. The sciences in general have attained uniformity in the logographic part of their writing. Why not extend this to ordinary writing? Such extension would be possible if syntactic uniformity existed among IE languages, at least among closely related IE languages. But such uniformity does not exist among these languages. That it does not exist is well illustrated by Bloomfield's example of the different syntactic and lexical structures of four languages for logographic 91 (ninety one in English, one and ninety in German, one and half five-times in Danish, four twenties eleven in French).

Uniformity in syntactic and lexical structure among the different languages of the Han Chinese family is sufficiently great to permit any given writing in the alphabet-included logographic system to be read in the same order, though with different pronunciations, depending on the particular language of the Chinese reader. This is what really lies back of the unique survival of Chinese character writing over four thousand years of use. This also gives good reason and good hope for the continuation of

Chinese character writing in the face of recent competition with a soon-to-be-coexisting alphabet of the self-sufficient type which reflects the different pronunciations among the different Han Chinese languages, and accordingly draws attention to the language barriers rather than bridging them. How well- or poorly - Chinese character writing bridges the language barriers among Han Chinese languages is another question.

Related to this question is the relative educational cost of learning or teaching a self-sufficient alphabet (in which every Han Chinese language may be written) versus the cost of Chinese character writing (in which every Han Chinese language is written). If something like complete control in reading Chinese can be had from recognizing 3000 to 5000 characters, and if reading with the help of a dictionary is possible after learning 26 hundred characters, the relative effort (from a western point of view) in learning to read Chinese and learning to read English is greater than a hundred to one. But to learn to write or read something directly as morphemes (in a composite character with sound specification) is quite different than learning the sound specification of a self-sufficient alphabet. To be literate in Chinese character writing is to be educated for two reasons (see 1.2 for the second reason). The first reason is that learning to read extensively is the same as learning an extensive vocabulary. Not so in English: learning the three R's, including writing, does not necessarily imply an educated individual.

There is an alphabetic order in our self-sufficient alphabet, which children often chant when learning their ABC's. But no Chinese child is taught to chant the sounds of thousands of characters. Instead, the modern Han Chinese child is presented with one or both of two pedagogic devices. Common vocabulary terms that the child has heard spoken - as those for mother, father, come, go - are written in Chinese characters on one side of a card, and are illustrated by picture drawn by cartoonists on the other side of the card. The second and more sophisticated device also uses pictures, but they are arranged facing Chinese characters which are simple (requiring few strokes at first) and then, progressively, opposite those which are more and more complex. Besides such modern pedagogic methods as these, there is a traditional lexicographic classification of Chinese characters, discussed above. Both the traditional and the modern classifications of characters help in learning (a) how to write Chinese, and (b) the lexical resources of Chinese; (a) and (b) are learned in one operation in China.

Karlgren conjured up a picture of different Han Chinese speakers reading one bulletin board in Shanghai. The Mandarin speaker and the Cantonese speaker stand side by side and they can not understand each other as they read aloud. But each can understand the single written message. To illustrate how this works specifically, we give one possible sentence in Chinese characters, and note that three rather than two spoken sentences are needed to show the closer relationship of Chinese character writing to Mandarin than to Cantonese.

His *home* *is* *not* *in* *America.*

This sentence consists of seven characters (written here left to right).

In reading aloud, the Mandarin speaker would say:

t'a^1 te^o čia^1 pu^1 cai^4 mei^3 kuo^2

he *of* *home* *not* *at* *beautiful* *country*

Later on, when out of sight of the writing, the Mandarin speaker might report what was written to his wife, repeating exactly what he had said when he read aloud.

In reading the same characters aloud, the Cantonese speaker would say:

t'a tek ka pat coi mei kuok

Later on, when out of sight of the writing, the Cantonese speaker, in reporting what was written to his wife, would not repeat exactly what he had said when he read aloud. Instead, when not directly influenced by the Mandarin selection in writing, the Cantonese husband might report:

k'öü ke ok k'ei m̥ hai mei kuok

It would be oversimplification to the point of falsification to say that the only spoken morpheme appearing unaltered in these three spoken sentences is mei^3 *beautiful* (from the description of America, which is *beautiful* *country*;) that the written character for *beautiful* is just pronounced like this (mei^3), and hence is especially appropriate in designating *America*, which in English also has m as its first consonant, as it does in Chinese; that Chinese in addition combines the sound with a symbol — a pleasant meaning (beautiful). The more realistic way of stating this takes cognizance of how mei^3 is combined with other morphemes in modern Chinese. Thus, mei^3 is often used as a transliteration character for syllables of the [me] type. As a morpheme, mei^3 is bound in most cases, and is less frequently used when the message beautiful is intended than is the compound /hau^3 k'an/ good looking. Accordingly, /mei^3 kuo^2/ scarely evokes the connotation of beautiful country when said or heard.

1.2. All scholars in old China, and most scholars in modern China would expand, by one step, our preliminary identification (1.1, above) of what it takes to be educated in Chinese culture. 'To be literate in Chinese character writing' is only the first step — a step which has to be taken in order to learn the contemporary classical language known as Wenyen (= Wenli), as well as to read Han Chinese; hence Chinese character writing serves a dual function.

An educated Chinese, then, uses character writing not only for writing and reading his own native language, but also for reading and writing Wenyen — in effect an additional language. This is one aspect of multilingualism that continues in modern China. The other aspect of modern multilingualism is the increased spread of 'acquired Mandarin' for the vast majority of speakers — namely, those whose mother tongue is any other dialect than Peking Mandarin. Thus, a person born and educated extensively in Shanghai will (a) speak Wu as his mother tongue, (b) be able to make himself understood, with some help from character writing, in his own variant 'acquired Mandarin' (1.3), and (c) be able to use the same character writing for enjoying Wenyen.

There are important differences between Wenyen as a contemporary, classical or literary language in respect to any Han Chinese language, and Greek or Latin as remote classical languages in respect to any other European language. To form scientific neologisms in English (e.g. 'cybernetics'), we generally borrow from the lexical resources of one or both of our remote classical languages, but do not expect educated people in general — much less literary artists — to begin to control them as do specialists ('classical scholars'); Shakespeare was said to know little Latin and less Greek. Educated people who are literate in any Han Chinese language have become so partly in order to be able to read and control the massive corpus of Wenyen which encapsulates, with the utmost terseness, a corpus of wisdom, as well as a lexicon that both overlaps and supplements dictionaries in the Han Chinese languages, and a structure that differs from all of them, but more from Mandarin, for example, then from Cantonese. Those who do not read are not entirely excluded from the Wenyen literature. Blind beggars can chant or quote such fragments as they have heard and memorized. Though speakers engage in conversations only in one of the Han Chinese languages, they may interlard an expression from Wenyen; or compose such expressions, for the sake of their brevity, in official government and legal documents, in business letters, often in private letters and in some news dispatches, and in virtually all newspaper advertisements and telegraph messages.

In contrast to the flamboyant style in which many of our advertisements are written, the use of Wenyen enables an advertiser to save space, or say more in the same space. Many words which require two characters in Han Chinese can be written with one character in Wenyen.

The cryptic effect of saving time and effort (and cost per word) that comes from ellipses of words in our telegraphic style is not entirely confined to telegrams in Western culture. It has been extended to writing some parts of botanical reports, for example. And the anthropologist Kroeber once experimented in publishing one of the California ethnographic surveys in telegraphic style. But such extension beyond telegrams seems to lack elegance, if not grammaticalness; Kroeber, for example, had to abandon his experiment in subsequent survey reports.

Chinese telegrams written in Wenyen achieve (a) greater compactness than is obtained in our telegraphic style, and (b) bear the flavor of an elegant literary tradition.

There are associations other than literary elegance which are conducive to the use of Wenyen as an on-going contemporary classic. Political mottos and slogans in Wenyen used to have an aura of being more worthy of respect than they would have been if they were given in any Han Chinese language. Sun Yat-sen's political mottos were in Wenyen. For any motto to be effective, the shorter the better, if the connotation is precise; this condition is met by Wenyen, but also by contracted Han Chinese. Since such leaders of

Communist China as Mao and Chou En-lai are counted as scholars, they would be able to quote from Wenyen. That one cannot be a good orator without quoting from Wenyen literature reflects an obsolete cultural value; knowledge of Wenyen is still respected, though exhibition of this knowledge has been said to have become politically inadvisable. Nevertheless, it is known that Mao is able to write elegant Wenyen poems; and that he still does write them.

About a generation ago, the serious study of Wenyen was begun by children still in grade school (e.g. in the seventh grade). Today Wenyen is studied by advanced students in universities. Whether or not this will transform Wenyen from a contemporary classic to a remote classic (like our Latin and Greek) must remain an open question.

The literary legacy from Confucius (6th century B.C.) to the end of the Ching dynasty (1644-1912) was written exclusively in Wenyen. Modern Chinese scholars learn to compose in Wenyen after the model of writers whose style differed from age to age. But in the Chinese apperception this corpus is so unified in structure and lexical selection that two instances from it — as widely separated historically as a century ago and over two millenia ago — are regarded as equally representative of Wenyen.

The structure of Wenyen and of Han Chinese in general is sufficiently different that the two would not be confused. For example, one of our Chinese students was writing to her mother about clothing; she was asked whether she could save space by using only one syllable (either yi^1 - or - fu^2) — as in Wenyen — or whether she would write both as a redundant compound for *clothing* (yi^1 - fu^2). Her response was that her mother would be shocked if she should write one character (one syllable, either yi^1 or fu^2) for *clothing*; her mother would say something like, "Oh, my daughter has lost her politeness and her education." She might think that her daughter had forgotten how to say yi^1 - fu^2, and was writing one or another because she was talking that way — ungrammatical in Han Chinese, but grammatical in Wenyen. The mother would never suppose that her daughter was attempting to write in Wenyen, since the rest of the letter in question was written in Mandarin.

Derk Bodde says it is necessary to "write only the single character *yi*, meaning 'clothing', or the single character *fu*, meaning the same" — but this holds for Wenyen, not for writing in any one of the Han Chinese languages. China's cultural tradition becomes blurred when the distinction between the Wenyen literary language and Han Chinese is not made: "The spoken language avoids such ambiguities ['syllables pronounced identically but having different meanings...'], however, by combining two syllables, similar in meaning, into a dissyllabic compound, and letting the compound stand for the idea which is common to both... *yi* and *fu*, for example, both mean 'clothing' (in addition to many other things); hence the spoken word for 'clothing' becomes *yi-fu*, a compound which will not readily be confused by the ear for any other compound...it is necessary to write only the single character *yi*, meaning 'clothing', or the single character *fu*, meaning the same, but not both of them put together (unless, of course, one were to write exactly as one would speak, but this, as we shall see, was not customary in the old China)." But it is customary in modern China, to write redundant compounds like yi^1 - fu^2 for clothing- clothing as one speaks them, unless one is composing in the Wenyen the literary language which differs from the language of any Chinese speech community. Since 1917 when Hu Shih persuaded people to write in plain talk (pai hua), the language of the community, most scientific writing, editorials, novels, dramas and love songs have been composed in written vernacular (1. 3).

1. 3. The double linguistic directive of Communist China amounts to this: (a) write as you speak (i. e. not in Wenyen); (b) speak Mandarin ('acquired Mandarin', if your mother tongue is some other Han Chinese language). The two parts of this directive are closely related; it would not do to say generally (a) 'write as you speak', if you speak only Cantonese; nor is it enough to say (b) 'speak Mandarin', for there are three or four different major groups of Mandarin dialects — though these are, for the most part, mutually intelligible. Hence the expanded directive for (b) does not include all of the Mandarin language but only that part of Mandarin called / p'u^3 t'un^1 hua^4/ common language (p'u^3 t'un^1 hua^4 is synonymous with kuo^2 u^3 national language).

When so restricted, the common national language turns out to be a mother tongue for the relatively few (a slight variant of how people speak who are brought up in Peking), while the vast majority face the task of learning 'acquired Mandarin', even if they are speakers of non-Peking

Mandarin dialects. If they are native speakers of a language far away in the south of China, the aim is not to acquire just any Mandarin dialect: the model of 'acquired Mandarin' is always the Peking dialect of Mandarin. Speakers who already know some other dialect of Mandarin would more closely approximate this model to begin with and would, therefore, have a much easier time acquiring 'acquired Mandarin' than would speakers of other Han Chinese languages.

Still, we know of no instance in which an educated speaker of one of the other languages has failed in an attempt to learn 'acquired Mandarin' to the point of becoming more or less intelligible to a Peking Mandarin speaker (though rarely without showing traces of his non-Mandarin native language, or even of his non-Peking dialect of Mandarin). The speaker of 'acquired Mandarin', if educated, can use character writing that he and the Peking Mandarin speaker share as a crutch in communication; either by writing in sand, in dust, on paper or on the blackboard; or by tracing the characters on the palm of one hand with the finger of the other; or by tracing the characters on a table with a chopstick dipped in soup; or even by tracing the characters in thin air.

In the second half of the 20th century, the speakers of a half dozen different languages in Mainland China are again faced with accepting — or rejecting — virtually the same politically sanctioned Mandarin dialect that was in political favor in the last dynasty of China (1644-1912). The foreign Manchus (Mongols) required officials from all the provinces to report to them in the Peking dialect, which was then known to the common people as language of the officials (kuan^1 hua^4). It functioned then as a lingua franca, while today there is apparent intent to have it replace the other Han Chinese languages. This may be regarded as merely programmatic, but it seems probable that no cultural shock would be felt in modern China by official insistence on the use of one dialect of one language — whether called a common language or national language (or as formerly, language of the officials); and it seems equally probable that a culture which has had much experience in official pressure of this kind has also learned how to preserve its languages and its literature.

A politically sanctioned dialect of Mandarin has for long been one of four supports of the ethnolinguistic monolith here called Han Chinese. Of two older supports of this ethnolinguistic monolith, one (Chinese character writing) is now competing with a self-sufficient alphabet, while the other (Wenyen) is being transformed from a contemporary classic to a remote classic. In comparison with 'acquired Mandarin', character writing, and Wenyen, the fourth support of the ethnolinguistic monolith is the strongest of all — the syntactic uniformity which lies behind the phonological and lexical diversity in a half dozen different languages, each made up of mutually intelligible dialects.

Chinese who are able to read the Wenyen literature with understanding (1.2, above) have necessarily mastered Chinese character writing (1.1, above) since the corpus of Wenyen is written exclusively in characters. Such Wenyen, when read aloud, is pronounced as the characters are pronounced by a speaker in his own language today, and no attempt is made to reconstruct former speech. Accordingly, the same Wenyen passage is pronounced differently by a speaker from Shanghai than by one from Peking. A given Wenyen passage may be intelligible in written form (character writing) but not intelligible in spoken form. If such a passage is unintelligible when pronounced, it would also be unintelligible if the pronunciation were written alphabetically (since all that alphabetic writing does is to specify sounds of pronunciation). To be intelligible, such a passage in Wenyen has to be either (a) written in character writing, or (b) translated from Wenyen into Wu or Mandarin, after which it can be written in alphabetic writing and still remain intelligible. Alphabetic writing and character writing are both possible for Wu, Mandarin, and other Han Chinese languages. Alphabetic writing is also suitable for Wenyen. Two Belgian Jesuits in their Romanisation Interdialectique, based on Karlgren's reconstructed Ancient Chinese, set up an alphabet (26 letters) which could write Wenyen or any Han Chinese language. If character writing were wholly given up in favor of alphabetic writing, Wenyen could still function as a contemporary classic language. Nor would Wu, Cantonese, Mandarin and the others suffer if written in romanization rather than in characters. Each symbol of the latter (character) is isomorphic with an alphabetic symbol sequence (syllable), by and large. However, since the shapes of the syllables are sufficiently different among the different Han Chinese languages to constitute a language barrier between them, it is necessary to say which

Chinese language is being written when the writing is in any romanization (or any other adaptation of the self-sufficient alphabet).

It is also necessary to distinguish between strings of syllables which represent written vernacular, and those which represent the actual talking of a Han Chinese language — talking not under the influence of Wenyen character writing or western educated bilinguals. The use of romanization is often but erroneously thought to represent exactly the way a Chinese language is spoken. The effective innovator of modern Chinese writing, Hu Shih, used to say that his writings in the vernacular were not, nor intended to be, in the form of actual speech ('only Chao writes the real spoken language'). The actual speech, the real spoken language is hereinafter called *spoken language*, and the vernacular represented in romanization or in characters will be called *written vernacular*.

It is interesting to note that Hu Shih played up the use of /pai hua/, but played down the use of romanization. Because Hu Shih expected to write /pai hua/ in characters for some decades, he felt no urgency in facing the knotty problems involved in romanizing Chinese languages and dialects.

1.4. In 'acquired Mandarin', as identified above, the written vernacular gives an overall impression of being somewhat more contracted — of utilizing fewer and sometimes different syllables — than the spoken language. Both share the same general structure; neither permits utterances quite as terse as those possible in Wenyen which is structurally more different from either written or spoken Mandarin than Mandarin is from other modern Chinese languages.

A speaker of 'acquired Mandarin' might say,

I am writing this letter in a slap-dash manner.

(1) uo^3 ma^2 ma^0 hu^3 hu^0 te^0 šie^3 če^4

[I]$_a$ [(horse ± horse) (tiger ± tiger) -ly]$_b$ [write]$_c$ [this

feŋ^1 šin^4

classifier letter]$_d$.

The literal translation of all sentences, here and below, matches one English word (the gloss) for each Chinese morpheme — and Chinese is said to be monosyllabic because most morphemes are exactly one syllable in length. Thus, one phrase, above, includes one morpheme šie^3 which is glossed as [write]. To mark the past, an additional morpheme would be included in the phrase, šie^3 le^0, and the phrase would be glossed [write perfective], reflected in free translation as I have written etc. In a compound some monosyllabic morphemes have literal glosses which only hint at or imply the referent for the compound as a whole. Others are not paradoxical, are quite explicit and even redundant, and hence show a better semantic fit with the meaning of the compound as a whole. The corresponding literal glosses of members of a compound are enclosed in parentheses. So also are the glosses of the repeated morphemes in reduplication. And a combined compound and reduplication is also enclosed in parentheses, as in sentence (1), where the literal glosses are (horse ± horse) (tiger ± tiger); the referent for this reduplicated compound is explicitly stated in the free translation which precedes sentence (1): in a slap-dash manner.

Though the speaker would certainly say sentence (1) in spoken language- and be always understood by anyone who comprehends any Mandarin dialect- he would be more apt to substitute (confused - cursive) for (horse ± horse) (tiger + tiger) in written vernacular, as is shown in sentence (2),

I am writing this letter carelessly.

(2) uo^{3} liau^{2} c' au^{3} te^{0} sie^{3} če^{4} feŋ^{1} šin^{4}

[I]$_{a}$ [(confused-cursive) - ly]$_{b}$ [write]$_{c}$ [this classifier letter]$_{d}$

Sentence (1) is uttered in ten syllables; sentence (2) is written in eight. The single difference lies in the spoken language selection of a reduplicated compound for (1) — (horse ± horse) (tiger ± tiger); and in the written language selection of a simple or non-reduplicated compound for (2) — (confused-cursive). But, of course, sentence (1) can be written (as indeed is apparent, above); and sentence (2) can be uttered. But sentence (2), when uttered, would be less widely understood by Mandarin speakers than sentence (1).

The first use of enclosing one or more glosses in brackets (and labelling each bracket with a subscript letter), is to permit comparison between closely similar sentences. Thus, sentences (1) and (2), above, are identical in brackets a, c, d. They differ only in bracket b.

The second use of enclosing one or more glosses in brackets is that it permits easy exposition of where the corresponding Chinese groups of morphemes are reorderable either in non-contrastive transformations of the same message, or contrastive transformations which change the message. There is an old stereotyped explanation, now generally abandoned, according to which some languages have more grammar than others. Languages which permit free reordering of any units, as in the non-contrastive syntax of Chinese, were supposed to have less grammar than, say, inflectional languages which have fixed limitations on the distribution of affixes flanking stems in word morphology. This stereotyped explanation was abandoned once the 20th century notion of levels within the grammar of a particular language was developed. In general, it was then clear that languages might differ in being grammatically more limiting at one level than another, rather than having more grammar in the whole of one language and less grammar in the whole of another language. But the older view of Chinese can be restated in terms of the new recognition of levels: Chinese shows more grammar (limitations on distributions) at some levels, and less grammar at other levels (hence more freedom, or choice, or possibility of reordering without contrast). Since brackets as units are often reorderable (i.e. permutable), without change in message, it is at this permutant level that restrictions on distribution are least in evidence. Thus in sentence (3), which follows, there are a half-dozen brackets; if now we hold the first bracket and the last bracket in constant position, it is possible to vary the positions of the four intervening brackets, labelled with subscripts from a to d, with the general message remaining constant.

He has never played happily at home with his own children.

(3) t'a^{1} pu^{1} c'eŋ^{2} kau^{1} kau^{1} šiŋ^{4} šiŋ^{4} te^{0} čai^{4} čia^{1}

[he]$_{a}$ [not ever]$_{b}$ [(high ± high)-(spirit ± spirit) -ly]$_{c}$ [at home

čuŋ^{1} ken^{1} hai^{2} cɨ^{0} men^{0} uan^{2} šua^{3}

middle]$_{d}$ [with (child-formative) plural]$_{e}$ [(play juggle)]$_{f}$.

The referent for the reduplicated compound of bracket c is happy; the ordering of the intervening brackets is $\overset{a}{[\text{he}]}$ b c d e $\overset{f}{[(\text{played})]}$. There are some half dozen possible permutations in the reordering of the four intervening brackets, namely:

(3.1) $\overset{a}{[\text{he}]}$ b d c e $\overset{f}{[(\text{played})]}$

(3.2) [he] b e c d [(played)]

(3.3) [he] d b c e [(played)]

(3.4) [he] d e b c [(played)]

(3.5) [he] e b c d [(played)]

(3.6) [he] e d b c [(played)]

(3.7) [he] b e d c [(played)]

Though these reorderings are possible, they are not selected at random, nor arbitrarily, nor by whim of the speaker. They exemplify the principle that the minimum syntactic span is more than one sentence in length, for the orderings in (3.1) to (3.7) are all response sentences, i.e., sentences uttered in response to a stimulus sentence which has its intervening brackets in an order parallel to that given in the response sentence. The general principle is: Say unto others as you have heard others say to you. This is an extension of Chao's rule of question-answer word [i.e., bracket] order: 'Ask as you would be answered.' That is, whatever the ordering of brackets in a question (or other kind of stimulus sentence), a parallel order will be given in the answer (or other kind of response sentence).

Thus, if someone says (stimulus sentence — here given in terms of its glosses in brackets), $[\underline{\text{he}}]_a$ $[\underline{\text{has}}\ \underline{\text{always}}]_b$ $[\underline{\text{at}}\ (\underline{\text{school}})\ \underline{\text{in}}]_d$ $[(\underline{\text{happy}})\text{-}\underline{\text{ly}}]_c$ $[\underline{\text{with}}\ (\underline{\text{student}})\ \underline{\text{plural}}]_e$ $[(\underline{\text{played}})]_f$; you select the b d c e ordering of sentence (3) brackets (response sentence (3.1)): $[\underline{\text{he}}]_a$ $[\underline{\text{not}}\ \underline{\text{ever}}]_b$ $[\underline{\text{at}}\ \underline{\text{home}}\ \underline{\text{middle}}]_d$ $[(\underline{\text{happy}})\text{-}\underline{\text{ly}}]_c$ $[\underline{\text{with}}\ (\underline{\text{child}})\ \underline{\text{plural}}]_e$ $[(\underline{\text{played}})]_f$. However, if someone says (stimulus sentence), $[\underline{\text{he}}]_a$ $[\underline{\text{not}}\ \underline{\text{ever}}]_b$ $[\underline{\text{with}}\ (\underline{\text{wife}})]_e$ $[\underline{\text{at}}\ \underline{\text{home}}\ \underline{\text{middle}}]_d$ $[(\underline{\text{happy}})\text{-}\underline{\text{ly}}]_c$ $[(\underline{\text{chats}})]_f$, you select the parallel b e d c ordering of sentence (3) brackets (response sentence (3.7)): $[\underline{\text{he}}]_a$ $[\underline{\text{not}}\ \underline{\text{ever}}]_b$ $[\underline{\text{with}}\ (\underline{\text{child}})\ \underline{\text{plural}}]_e$ $[\underline{\text{at}}\ \underline{\text{home}}\ \underline{\text{middle}}]_d$ $[(\underline{\text{happy}})\text{-}\underline{\text{ly}}]_c$ $[(\underline{\text{played}})]_e$. And so on, for the other response sentences with still different orderings of intervening brackets. Accordingly, the possibility of reordering without change of denotation in the message appears to be not entirely free, since it — and stress and intonation as well — parallels the choice in ordering of the preceding speaker. But it is relatively free. Many alternatives are possible; but only one alternative in the response — the alternate which echoes order, stress, and intonation of the stimulus — is polite as well as correct.

The possibility of reordering brackets with change of message also exists. This second kind of reordering is relatively limited, and its restrictions are not influenced by the parallel syntax of the preceding speaker. It is in fact much like functional word order in English (e.g., [*tiger*] is subject, or topic before the predicate or comment [*killed* *the* *man*], while [*man*] is subject or topic when preceding the predicate or comment [*killed* *the* *tiger*]). But *topic* is not always translatable as English subject, and *comment* — unlike English predicate — may include embedded *topics*. Also, the *topic*-*comment* level is explicable in a syntactic span that is longer than one sentence. We return to the *topic*-*comment* level after considering lower levels — shorter spans of morphemes within the brackets.

It is the bracket as a whole that is reorderable and, therefore, sometimes analogous to a 'linguistic word' in languages with a 'word' level. The single morphemes within a bracket, however, might be said to represent 'ethnolinguistic words', since children are taught to learn and to write single morphemes — separate characters in the traditional writing system — and adults pay for telegrams according to the number of single morphemes rather than according to the number of brackets in a message. There is a Chinese label for the single morpheme (equivalent to the single character) — cɨ^4 — but no label for the bracket, our reorderable unit. The reorderable bracket and the single morpheme coincide when a bracket includes just one morpheme, as in the first bracket of each sentence given above, and as in bracket $_c$ in sentences (1) and (2). All other brackets in sentences (1), (2) and (3) include more than one morpheme.

Literal glosses are given for the single morphemes, even when in parentheses; but the referent or connotation of the reduplication and/or compound in parentheses is given in free translation, as in the last bracket in sentence (3): [(*play juggle*)] meaning *play* or *played*.

In text or discourse frequency, as would be expected, the morphemes that always function as lexical or major morphemes (capital M, in our notation) are most conspicuous. Scarcely more than one in five — and fewer in inventory count than in text frequency — function as grammatical or minor morphemes (small m, in our notation), analogous to functive words and affixes (suffixes or prefixes or infixes) in languages with a word level. But, most surprising of all to Western eyes, a very large fraction of single morphemes function alternatively as minor morphemes in some brackets and as major morphemes in other brackets ($\frac{m}{M} \sim \frac{M}{m}$, in our notation, as exemplified in sentence (4), below).

It becomes immediately clear, when applied to English examples, how the notation works when lexical or major morphemes are represented by M, when morphemes in the grammatical inventory (mostly suffixes) are represented by m, and when morphemes which function alternatively as either — but in different sequences — are marked by $\frac{M}{m} \sim \frac{m}{M}$. Thus English *repeatedly* represents a sequence M-m-m, since a lexical or major morpheme, *repeat* (M) is followed by -m-m, two minor morphemes which are always suffixes,

-*ed* -*ly*. 'Once a suffix always a suffix' is a rule which holds generally for English, but some few suffixes (minor morphemes) appear also as lexical or major morphemes (or vice versa). Notation used for instances of such overlap, is $\frac{m}{M}$ when the morpheme is functioning as a suffix, but $\frac{M}{m}$ when the same morpheme is functioning as a lexical or major morpheme. Thus, in *full of beans* and *less than half*, both *full* and *less* are functioning as major morphemes. However, we represent both as $\frac{M}{m}$ because both *full* and *less* may also function as minor morphemes, as in *carefully*, *carelessly*. Here -*full*- and -*less*- are represented as $\frac{m}{M}$ and the sequence as a whole is represented as M-$\frac{m}{M}$-m. Occasional parallel instances of $\frac{m}{M} = \frac{M}{m}$ are found when a suffix or sequences of suffixes are uttered as words, generally as technical words — *ism* or *isms* from *socialism* and the like, *emic* and *etic* from *phonemic*, *phonetic* and the like.

The rule, 'once a suffix always a suffix' is a general rule for English, but not for Chinese. The total inventory for minor morphemes — m's which always function as grammatical or minor morphemes — is surprisingly small. The total inventory of major morphemes — M's which always function as major morphemes — forms the largest fraction of lexical morphemes in the dictionary. But hundreds of morphemes belong equally to (a) the grammar — $\frac{m}{M}$ when they serve a grammatical function; and to (b) the dictionary — — when they function like other lexical or major morphemes — that is, like the majority of M's.

It is only a Western bias to say 'a language should not be confused with its dictionary' or that ' the dictionary of a language is an appendix to its grammar'. In the Chinese view, all instances of $\frac{m}{M}$ - $\frac{M}{m}$ belong simultaneously to the grammar and the dictionary. And instances of $\frac{m}{M}$ - $\frac{M}{m}$ are numerous and highly recurrent in any text. Instances of m's (minor or grammatical morphemes which keep their grammatical function in all occurrences) are highly recurrent, to be sure, but not numerous in inventory count. So far as our sample is predictive for 'acquired Mandarin', it suggests that almost half of t occurrences of morphemes in texts belong simultaneously to the grammar or minor morpheme inventory and to the major morpheme inventory ($\frac{M}{m}$), where the meaning is generally altered.

Thus, feŋ^1 is glossed as classifier relevant to envelopes or small packages, as in sentences (1) and (2) above, and (4a) below, when it functions as $\frac{m}{M}$. But when functioning as $\frac{M}{m}$, feŋ^1 is glossed as seal or sealer as in sentence (4),

He has sealed that letter.

(4a)	t'a^1	feŋ^1	le^0	nei^4	feŋ^0	šin^4
	he	seal	perfective	that	classifier	letter
	$[M]_a$	$[\frac{M}{m}$	$m]_b$	$[\frac{m}{M}$	$\frac{m}{M}$	$M]_c$

The envelope is on the desk.

(4b)	šin^4	feŋ^1	cai^4	čuo^1	šaŋ^0
	(letter	sealer)	at	table	on
	[M	$\frac{M}{m}$]	$[\frac{M}{m}$	M	$\frac{m}{M}]$

Here and in the following sentences the glosses for compounds and for reduplication are enclosed in parentheses, but otherwise all notation (including that for brackets, M, m, and $\frac{M}{m} \sim \frac{m}{M}$) is placed in the third line under the uncluttered Chinese morphemes and their literal English translations (glosses); the free translation precedes the numbered sentences. The morpheme feŋ^1 appears as the verb seal and as the classifier for letter in sentence (4a); it also appears in the compound for envelope (šin^4 feŋ^1) (letter sealer) in sentence (4b). The compound including šin^4 does not occur with feŋ^1 as classifier.

Simple compounds are represented as (M M) when each member can also appear as a single major morpheme M, not in compound. The last bracket in sentence (3) provides an instance of this: (play juggle), where two different free major morphemes denote play. They are free because each M may appear out of compound, uan^2 to play, šua^3 to juggle.

The first or second member of the simple compound, or even both members, may be $\frac{M}{m}$ instead of M (see above), as in the first bracket of sentence (4b) where (letter sealer) means envelope — an instance of (M $\frac{M}{m}$).

When the first member of a compound is a bound M- or the second is a bound -M, or when both members are bound, the simple compound is represented as hyphenated (M-M). Thus, the M- for confused (liau^2) is bound while the M for cursive (c'au^3) is free in the simple compound (M-M): (confused - cursive) for careless in sentence (2).

Whether the simple compound is of the (M M) or (M-M) type, the relationship of the literal glosses to what the compound denotes as a whole may be either self-exlanatory or even redundant, on the one hand, or paradoxical in the sense that the literal glosses, especially when coordinate, pose a paradox which is resolved in the meaning of the compound as a whole. Thus, the reduplicated compound in sentence (1) may be reduced to a simple (M M) compound, (horse tiger), meaning carefree, not particular, not fussy. Other simple compounds with coordinate members are more or less like this: (east west) meaning something (tuŋ^{1} ši^{0}); and (horizontal vertical) meaning any way (heŋ^{2} šu^{4}); and (come go) meaning checking account (lai^{2} uaŋ^{3}). So also, (to beat to sort) meaning to clean up (ta^{2} lü^{3}); this is an instance of (M M) which may be shortened to the second member which alone is less specific than the compound (lü^{3} to sort, to put in order). Some coordinate compounds are self-explanatory in the sense that the meaning of the whole may be inferred from the meaning of the members a (to repair to sort) for to fix up (šiu^{1} li^{3}); this instance of (M M) may be shortened to — or expanded from — the first member of the compound (šiu^{1} to repair, to mend). Examples of the self-explanatory kinds of simple compounds have appeared in the preceding sentences: (confused-cursive) meaning careless; (play juggle) meaning play; (letter sealer) meaning envelope. In some of these self-explanatory compounds, the first member is subordinate to the second. In some, the glosses of the two members overlap in meaning, and then the two members are again coordinate. Thus, redundantly, (clothing - clothing) means clothing (i^{1} fu^{0}); and (body - body) means body (šen^{1} t'i^{3}); and (friend -friend) means friend (peŋ^{2} iou^{0}); and (mad raving) means insane feŋ^{1} k'uaŋ^{2}); and (look see) means to see (k'an^{4} čien^{0}). If the first member is modifier in respect to the second member, the compound may be either paradoxical or self-explanatory as (move thing) meaning animal (tuŋ^{4} u^{4}); as (ground diagram) meaning map (ti^{4} t'u^{2}). Compare (offending person) for criminal, and (interior person) for wife (of speaker), where the second member is the same — (fan^{4} ren^{0}) and (nei^{4} ren^{0}) respectively. So also, for the next pair, (fierce hand) means murderer (šioŋ^{1} šou^{0}) and (rifle hand) means a good shot, an expert marksman (c'iaŋ^{1} šou^{0}). For the last compound, reversal of order is possible: (hand rifle) for revolver (šou^{3} čiaŋ^{1}). In fact, this is the only compound cited so far in which the order of (M_1 M_2) can be changed to (M_2 M_1), and still remain a compound. In (new smell) for news the second member is glossed for its value in spoken language, but is closer to the meaning of the whole compound if it is glossed for its Wenyen value (new hear) for news (šin^{1} uen^{2}). If the speaker does not know Wenyen, he has to learn this compound without the crutch of analysis; so also, for the following compound whose members can be interchanged without change of meaning: either (hear see) or (see hear) for observe, observation (perception by ears and eyes, or by eyes and ears, but not by nose) — (uen^{2} čien^{0}) ~ (čien^{4} uen^{2}) — even though uen^{2} out of compound now means to sniff, to smell. Finally, we list compounds in which the second member is related to the first as its object, as (flow blood) for to bleed (lieu^{2} šüe^{3}); as (move-body) for to start on a journey (tuŋ^{4} šen^{1}); and, more paradoxically, as (use-merit) for to study hard (ioŋ^{4} kuŋ^{1}).

In the preceding two-member compounds, the gloss selected for each member was influenced by the meaning that the morpheme in question would mark when out of compound. If the meaning that the morpheme marks when in compound were selected, then instead of (fierce hand) for murderer, we would gloss (crime hand), instead of (new smell) for news, (new hear); uen^2 is fairly active in compounds.

Triple compounds are marginal in inventory, or seem slightly out of pattern when used in some neologisms, but suggest no uncertainty or ambiguity in analysis of the immediate constituent type. A triple compound is represented as (M M) (M) or as (M) (M M) when the members are free major morphemes, as in (season wait) (wind) for the monsoons (č'i^4 hou^4) (feŋ^1), and as in (yellow hot) (sickness) for yellow fever (huaŋ^2 re^4) (piŋ^4); (yellow hot) does not occur in sequence except before (sickness). It is possible to distinguish between a simple compound preceded by the major morpheme for yellow in sequence with minor morpheme, and a triple compound in which M for yellow is the first member: (M-m) (M M) represents (yellow-of) (cover car) for yellow rented car [huaŋ^2 te^0 (pau^1 č'e^1)]. The same three major morphemes in triple compound — (M) (MM) but without the minor morpheme te^0 — appear in 'acquired Mandarin', perhaps under Shanghai influence, as (yellow) (cover car), meaning ricksha [(huaŋ^2) (pau^1 č'e^1)]. When one or more members of a triple compound are bound, this is represented by hyphens, as (M - M) (- M) for (descend-fall) (-squadron) meaning paratrooper [(čiaŋ^4 luo^4) (tuei^4)]; compare (descend-fall) (umbrella), meaning *parachute* [(čiaŋ^4 luo^4) (san^3)]; and the sequence (*cold-hot*) which occurs only in sequence (*cold-hot*) (*tabulation*) meaning *thermometer* [(han^2 ǯu^2) (piau^3)]. Perhaps (*not have-wire*) (*electricity*) meaning *radio* is an extended calque from *wireless* [(u^2 šien^4) (tien^4)]. It is not certain whether the following example represents a triple compound (M) (M M), or a phrase in which the first member, glossed *to send out*, precedes an (M M) compound, [*to send out* (*spleen steam*)], meaning, in either analysis, *to lose one's temper* [(fa^1) (p'i^2 č'i^0)].

Sequences of paired compounds also occur (see below). So also do paired reduplications — i.e. reduplicated compounds. Both are extensions of simple compounds and simple reduplication. Sequences of same morphemes in reduplication and their glosses are inclosed in parentheses, as are members of compounds, but the relation of the repeated members in a reduplication is shown by different notation than the relationship of different morphemes in a compound. For bound morphemes, the plus sign is used when the morpheme is repeated in reduplication (M + M); and, as already noted, the hyphen is used for connecting different bound morphemes in compound (M - M). Most common kinship terms are instances of (M + M), though they are reduced to single M in letter writing — in epistolary style — as substitute for first person, *I*, the writer's kinship term in respect to the person written to; and as substitute for second person, *you*, the addressee's kinship term in reference to the writer. But other than in letter writing (and in Wenyen), the majority of kinship terms appear as instances of (M + M) reduplication.

reduplication. Thus, in instances specifying younger siblings: (*brother* + *brother*) for *younger brother* (ti^4 ti^0), and (*sister* + *sister*) for *younger sister* (mei^4 mei^0). So also in instances specifying older siblings, (*sister* + *sister*) for *older sister* (čie^3 čie^0), and (*brother* + *brother*) for *older brother* (ke^1 ke^0). In all instances but the last, these M's appear not only in (M + M) reduplication but also in (M - M) compounds, as (*younger brother*-*younger sister*) meaning *younger siblings of either sex*, collectively (ti^4 mei^4), and as (*older sister*- *younger sister*) meaning *sisters irrespective of relative age*, collectively (čie^3 mei^4). On the other hand, though there is an (M - M) compound for (*older brother*-*younger brother*) meaning *brothers irrespective of relative age*, the first member of the compound is not selected from the (M + M) reduplication, as given above, though the second member is so selected — (šioŋ^1 ti^4) rather than *(ke^1 ti^4).

The relationship of the repeated members of a reduplication is indicated by the plus-and-minus sign when the repeated morpheme is free — that is, also occurs out of reduplication and out of compound, as well as in (M ± M) reduplication. Thus, as a free morpheme, t'ai^4 *too much* occurs before modifiers, including quantifiers (as *too much red*, *too many*); but in reduplication, represented as (M ± M), (*too much* ± *too much*) means *wife* (t'ai^4 t'ai^0). For morphemes which may enter into such optional reduplication, the difference in meaning between the single morpheme and the reduplicated morphemes, when there is a difference, is quite diverse — more or less paradoxical, as in the preceding example; also, having reference to command or entreaty or commiseration, as in the sentences immediately following; and also marking other notions, as distributiveness or liveliness. If the repeated M's in sentence (5) were reduced to a single M, the sentence would declare You are helping that man; as cited, with (M ± M), sentence (5) is imperative, You give that man a hand.

(5) ni^3 (paŋ^1 paŋ^0) nei^4 ke^0 ren^2
you (help ± help) that classifier man
$[M]_a$ $[(M \pm M)]_b$ $[\frac{m}{M}$ m $M]_c$

In sentence (6), the M for thin is out of reduplication, He's very thin.

(6) t'a^1 hen^3 šou^4
he very thin
[M] [m M]

In sentence (7), the reduplication adds the notion of commiseration, as in the free translation, He's rather thin (poor thing).

(7) t'a^1 (šou^4 šou^4) te^0
he thin ± thin -ly
[M] [(M ± M) -m]

But in a parallel reduplication, [he (fat ± fat) -ly], reduplication adds the notion of something jolly or lively, as when we say of a baby He's fat (cute). In sentence (8), the reduplication adds the notion of entreaty, Try and eat this to see (to see if it's sour or sweet or bitter or poison, or whatever in context).

Try and eat this.

(8) (č'ɨ^{1} č'ɨ^{0}) k'an^{0}

(eat ± eat) look

[(M ± M) $\frac{m}{M}$]

Compounds also appear in pairs, as do reduplications. An example of the former is represented by (M - M) - (M - M), as for example, (older brother- younger brother)-(older sister-younger sister) meaning all siblings (sioŋ^{1} ti^{4})-(čie^{3} mei^{4}). Examples of the latter are found in sentence (1) (horse ± horse)-(tiger ± tiger) meaning in a slap-dash manner; and in sentence (3) (high ± high)-(spirit ± spirit) meaning happy. The reduplicated compound marks a more lively meaning (e.g. slap-dash) that the simple compound (e.g. carefree), when there is a difference. The difference between sentence (3), with a reduplicated compound, and sentence (9), with simple compound, is a matter of what marks liveliness — marked by reduplicated compound in (3), and by a minor morpheme for very in (9), I'm very happy.

(9) uo^{3} hen^{3} (kau^{1} šiŋ^{4})

I very (high-spirit)

[M]$_{a}$ [m- (M - M)]$_{b}$

In all instances of compounds and reduplication, so far, the constituent members are represented by major morphemes (M's). But two-member compounds, or compound-like sequences of two members, also occur in which one member is a minor morpheme, m or $\frac{m}{M}$. The most productive of the latter are formative compounds.

In formative compounds the major morpheme is dependent. Thus, (*child*-*formative*) means *child* (xai^2 cɨ^0); and (*grandchild*-*formative*) means *grandchild* (suen^1 cɨ^0); and (*offspring*-*formative*) means *son* (er^2 cɨ^0). All of these represent instances of (M - $\frac{m}{M}$). But compare the last instance for *son* with instances of (M - $\frac{M}{m}$), as (*female offspring*) for *daughter* (nŭ^3 er^0), on the one hand, and with the members of this compound reversed, when the gloss for *offspring* is specialized to *male offspring*, as in (*male offspring*-*female*) meaning *offspring of either sex or both sexes* (er^2 nü^3).

When the first M in a compound is free, the *formative* represented above as $\frac{m}{M}$ often functions as $\frac{M}{m}$ rather than as a formative; indeed, though the same morpheme, it is glossed — after free M — as *potential* and represented as $\frac{M}{m}$ in a simple (M $\frac{M}{m}$) compound. Thus, (*melon potential*) means *melon seed* (kua^1 cɨ^3) in contrast to *melon* (kua^1); and (*fish potential*) means *roe* (ü^2 cɨ^3), beside *fish* (ü^2); and (*gun potential*) means *bullet* (č'iaŋ^1 cɨ^3), beside *gun* (č'iaŋ^1); and (*female potential*) means *woman* (nŭ cɨ^3). Similarly, when the members of the compound for woman — are reversed, the new compound, ($\frac{M}{m}$ - M), means *offspring of either sex or both sexes* (čɨ^2 nü^3).

Localizer compounds in which the localizer has full tone are less common than formative compounds. The localizer, as second member of a compound, is represented by $\frac{m}{M}$ in (M $\frac{m}{M}$): (*heaven below*) means *the world*, (t'ien^1 šia^4); (*city inside*) means *downtown* (č'eŋ^2 li^3); (*horse above*) means

2

immediately, at once (ma^3 šaŋ^4). But $\frac{m}{M}$ for localizer may also follow M out of compound, as in city in, reversing the prepositional order of English, in the city, and represented not as *[(M - $\frac{m}{M}$)] but as [M - $\frac{m}{M}$] — [č'eŋ^2 li^0]. (Compare the same morphemes as a compound for downtown, above.) Furthermore, morphemes here represented as $\frac{m}{M}$ alternate as $\frac{M}{m}$, when they are glossed somewhat differently. As $\frac{m}{M}$, šia^4 is glossed down but as $\frac{M}{m}$ šia^4 is glossed to descend; so also, li^3 is inside, within as $\frac{m}{M}$, but is lining of garment in the compound ($\frac{M}{m}$ - $\frac{m}{M}$) — (li^3 cɨ^0); so also šaŋ^4 is on, above as $\frac{m}{M}$, but is to go up, to ascend as $\frac{M}{m}$.

Within the bracket, the major morphemes, M and $\frac{M}{m}$, and the minor morphemes, m and $\frac{m}{M}$, which are free and not in parentheses, function as parts of speech. So also, the parentheses for compound or reduplication or reduplicated compound count as one part of speech, though the members of the compound or the repeated morphemes in reduplication are not so classified. It is only the part of speech of the whole parentheses that is relevant to the syntax of the phrase or of the sentence, though it would be possible to assign class membership to each member of the compound. Thus, the reduplication (too much-too much) means wife and counts as noun, irrespective of the form class of the repeated morpheme enclosed in parentheses (t'ai^4). The compounds and reduplications represented by parentheses as wholes — as well as major morphemes, M or $\frac{M}{m}$, out of parentheses — both distinguish four parts of speech: verb, noun, adjective, and adverb; and a fifth part of speech, pronoun, is distinguished by M alone (out of parentheses, since pro-

nouns are neither reduplicated nor enter into compounds).

Within a phrase bracket, the minor morphemes, m or $\frac{m}{M}$, give specification or support to the major parts of speech, or otherwise complement the major morphemes, M or $\frac{M}{m}$; and some minor morphemes also distinguish what might be called minor parts of speech: demonstrative, numeral, classifier, as well as localizer, directional and other complements.

In self-sufficient sentences, each phrase bracket includes at least one major morpheme, M or $\frac{M}{m}$. Stated conversely, no brackets have been cited, so far, with one or two minor morphemes, m or $\frac{m}{M}$, to the exclusion of major morphemes. Such phrase brackets do occasionally occur, however, but only in sentences which follow or are followed by a self-sufficient sentence in a longer discourse. This is one dimension of macro-syntax. When not self-sufficient, either the stimulus sentence or the response sentence may be linguistically dependent on a self-sufficient sentence in the longer discourse; or ethnolinguistically dependent on the context, as by pointing at something prior to beginning a sentence with a demonstrative functioning as pronoun (če^4...This...in sentence (14), below). Such dependence, whether linguistic or ethnolinguistic, is a clear indication of macro-syntax. A phrase bracket which includes minor morphemes (m or $\frac{m}{M}$) exclusively is sure to be in macro-syntax, while in the narrower sentence syntax, all phrase brackets include a major morpheme (M or $\frac{M}{m}$).

In sentences (1)$_d$ and (2)$_d$, [this classifier letter] represents a sequence of two minor parts of speech (demonstrative and classifier) followed

by a major part of speech (noun), in a phrase bracket — $[\overset{m}{M}\ \overset{m}{M}\ M]$ — which functions as a noun phrase. The same two minor parts of speech are followed by a noun compound in parentheses in the self-sufficient sentence,

This card is pretty.

(10)	če^4	čaŋ^1	k'a^3	p'ien^4	hau^3	k'an^4
	this	classifier	(card	sheet)	(good	look)
	$[\overset{m}{M}$	$\overset{m}{M}$	(M	$\overset{M}{m})]_a$	$[(\overset{M}{m}$	$M)]_b$

But though the same two minor parts of speech may be the only members of a bracket phrase, $[\overset{m}{M}\ \overset{m}{M}]$, without M for noun, the bracket as whole is a noun phrase, as in

This one is pretty.

(11)	če^4	čaŋ^1	hau^3	k'an^4
	this	classifier	(good	look)
	$[\overset{m}{M}$	$\overset{m}{M}]_a$	[(M	$M)]_b$

The difference between sentences (10) and (11) is that (10) is a self-sufficient sentence, while (11) is found only in macro-syntax, as defined above. The ethnolinguistic dimension of macro-syntax is well illustrated in the next sentence: standing before a collection of photographs or paintings, one might say,

Every one of them is pretty.

(12)	čaŋ^1		čaŋ^0	tou^1	hau^3	k'an^4
	(classifier		classifier)	all	(good	look)
	$[(\overset{m}{M}$	±	$\overset{m}{M})]_a$	[m	(M	$M)]_b$

Heard as an isolated sentence from the playback of a tape recording, one could not say from sentence (12) just what was pretty — it might be, on the one hand, a postcard, or photograph, or painting, or even sheets of blank paper, and it would be in this domain that one would first guess what reference was intended; on the other hand, since the particular classifier in sentence (12)$_a$ is also relevant to other objects with extended surfaces, including beds, desks or tables, stools, rugs, and linen sheets, one's first guess would have been wrong if the actual reference was to a collection of stools or rugs, for example. The compound classifier here is, generically, a distributive collection of objects with extended surfaces. But the same reduplicated classifier also occurs before a compound noun in the first bracket of the self-sufficient sentence,

All the cards are pretty.

(13)	$čaŋ^1$	$čaŋ^0$	$k'a^3$	$p'ien^4$	tou^1	hau^3	$k'an^4$
	(classifier	classifier)	(card	sheet)	all	(good	look)
	$[(\frac{m}{M}$	$\pm \frac{m}{M})$	(M	$\frac{M}{m})]_a$	[m	(M	$M)]_b$

All brackets which include a classifier function as noun phrase. Sentences (11) and (12) are in macro-syntax, while (10) and (13) are self-sufficient. In self-sufficient sentences, the order is either demonstrative or numeral before unreduplicated classifier before noun (10); or else reduplicated classifier before noun (13); see also sentence (61) for adjective phrase between classifier and noun. In macro-syntax, the order is either demonstrative or numeral before classifier, as in sentence (11); or else repeated classifier (reduplication) as in sentence (12), without following noun in either order, but with the bracket, nevertheless, functioning as noun phrase. Reduplication of the classifier is incompatible with prior demonstrative or numeral. Such restrictions are more closely associated with the classifier than with the noun (which may occur with or without classifier) or the demonstrative or numeral (which may also occur with or without classifier). But unreduplicated classifiers are flanked by following noun and preceding demonstrative or numeral.

Many classifiers may be transformed from a classifier (a minor part of speech) to a major part of speech (as noun or verb). Thus, čaŋ^1 is transformed from classifier ($\frac{m}{M}$, above) to verb ($\frac{M}{m}$), as in (čaŋ^1 k'ai^1) to open up. A better example is feŋ^1 which may be transformed from a verb, to seal ($\frac{M}{m}$), to a classifier ($\frac{m}{M}$), relevant to the domain of letters and small packages, as in sentence (4a). So similarly, pen^3 may be transformed from a noun (glossed in various ways — as cost, fiscal capital, and when in compounds, as same, origin, root, booklet) to classifier relevant to a narrow domain — to bound sheets of paper or cloth, as notebooks, magazines, pattern books, and ordinary books.

Beside the hundred or so classifiers which are transformable in this way (from $\frac{m}{M}$ to $\frac{M}{m}$), there are relatively few classifiers which are never so transformed. Hence, the latter are represented as non-transformable minor morpheme (m) rather than transformable $\frac{m}{M}$. Thus, čɨ^1 is relevant to the medium sized objects, animals, or a member of a pair; k'e^1 to the domain of plants and especially trees; tuo^3 to the domain of flowers; k'uai^4 is relevant to lumps or pieces, and may also be glossed *dollar*, but only in macro-syntax; ke^4 is relevant to the widest domain of all and is also substitutable as a general classifier in place of the more domain-restricted transformable classifiers. Still the domains to which ke^4 are relevant are not entirely unrestricted — ke^4 does not occur before nouns for sheets of paper, nor before nouns for long strips as river or thread, flowers and trees, etc., when preceded by numerals higher than *one*, though it may occur before such nouns when preceded by the numeral *one* or a demonstrative.

Sentence (14), following, shows a non-transformable minor morpheme in an affixal relationship to one of the major morphemes: M-m is glossed as *I-of* and means *my* before the noun compound (*heaven below*), translated freely as *world* in a sentence excerpted from a longer discourse, *This is my world*.

(14) če^4 sɨ^4 uo^3 te^0 t'ien^1 šia^4

this *is* *I* - *of* (*heaven* *below*)

$[\frac{M}{m}]_a$ $[M]_b$ $[M$ - m $(M$ $\frac{m}{M})]_c$

Minor morphemes affixed to major morphemes — as the instance represented by M-m in sentence (14) — are very low in inventory count, but occur over and over again in texts. Some few are suffixed (-m); still fewer are prefixed (m-); and an occasional one or two are infixed (-m-). An example of the latter appears in echo reduplication which is a variant of the simpler kinds of reduplication and compounds noted above. Thus, the simple (M - M) compound — (hu^2 t'u^2) glossed, as the members would be glossed out of compound, as (paste-smear) for the adjectival verb, blurred or muddled — is transformed from verb part of speech to adverb part of speech, meaning in a muddled manner, good and muddled, blurrish, when the simple compound is wholly reduplicated (both members repeated), and also when partly reduplicated (first member repeated, with intervening infix): (hu^2 hu^0) - (tu^2 tu^0); and also (with infix -li^0-), (hu^2 li^0 hu^2 tu^0); the latter is an instance of echo reduplication.

For an example of prefix for not alternating with verb for have no, compare (15), (16), and (17):

I don't have (as response sentence).

(15) uo^3 mei^2 iou$^{3(\sim 0)}$

I not have

$[M]_a \quad [\frac{m}{M} - M]_b$

I don't have money.

(16) uo^3 mei^2 iou^3 č'ien^2

I not have money

$[M]_a \quad [\frac{m}{M} - M]_b \quad [M]_c$

I haven't any money.

(17) uo^3 mei^2 č'ien^2

I have no money

$[M]_a \quad [\frac{M}{m} - M]_b$

But it is possible to reorder sentence (16) from a b c to c a b — as in

English, from I don't have money to Money I don't have. Such reordering is not possible in sentence (17) where mei^2 functions as a verb before a noun included in the same bracket because the noun in this kind of sequence is not independently permutable. (One can't end a sentence with mei^2.) The same noun is included in a separate bracket when it is reorderable — from last bracket, as in sentence (16) to first bracket in sentence (16): [čien^2] [uo^3] [mei^2 iou^3] Money I don't have.

In the western tradition of alphabetic writing, space serves as a word marker. If a word contains unusually many morphemes — say a prefix (m-), stem (M), and suffix (-m) — the whole m-M-m sequence is still set off by space as, for example, un-usual-ly in this sentence. In the Chinese tradition of character writing, space — or whatever Chinese use as the equivalent of space — serves to separate successive syllables, and that amounts to successive morphemes in almost every case. This tradition of writing permits recognition of characters (by the device of keeping each character together). Thus, separate morphemes by separate character tradition may be continued in modern alphabetic writing; indeed, it is reflected in the way the 'acquired Mandarin' is written here — by keeping morphemes apart by space, as in fei^1 č'aŋ^2 te^0 un-usual-ly. It is we who add the parentheses (to group together morphemes in reduplication or in compound), just as we add the brackets (for phrases which include one or more morphemes), and just as we insert -, +, ± as notation within parentheses for formulae (but not for the Chinese part). These are merely notational additions which still permit space to serve as a morpheme boundary marker, irrespective of whether a given morpheme is always a major morpheme (M); or always a minor morpheme (m); or sometimes one and sometimes the other ($\frac{M}{m} \sim \frac{m}{M}$).

There is no lack of clear-cut distinction between an M which functions as a major part of speech — say the pronoun for I (uo^3) — and an m which is always a minor morpheme of the kind that never alternates as a major morpheme. Every instance of the latter (m) is an affix — either a prefix or a suffix (or an infix, but there are relatively few infixes). But though an affix, the morpheme represented by m is still separated from the other morphemes by space. Thus, in the case of the M-m sequence for I-pluralizer, meaning we, there is a space between the major morpheme (uo^3) and the suffixed minor morpheme (men^0): uo^3 men^0 we.

This is consistent and useful. It is consistent because space separates morphemes, not words. It is useful for languages whose affixed (m's) may have either a narrow dependence range in which the dependence extends only to adjacent major morphemes (hence as part of one 'word' in languages like English with a word level), on the one hand; or on the other hand, whose affixes (m's) may just as characteristically have a wide dependence range in which the dependence extends to a whole phrase, clause, or sentence.

The wide dependence range is not altogether unknown in languages with word level like English, but it is not characteristic of them. It is found in such expressions as "the King of England's hat" or, with an even wider dependence range for the same suffix ('s), in the sentence that H. L. Mencken (1936) cites as a curiosity: "That umbrella is the young lady I go with's." As Y. R. Chao (1960) has shown, analogues of such wide dependence range for suffixes are not at all curious but instead just as characteristic of Chinese structure as the narrow dependence range exemplified above by [I-pluralizer] meaning we. A similar example is [I-of] meaning mine in a phrase without following noun, or meaning my whatever the noun is, if a noun is included in the phrase (uo^3 te^0, before noun). Compare now two suffixes in narrow range dependence: [I-pluralizer-of] meaning ours in a phrase without following noun, or else our (uo^3 men^0 te^0, before noun).

The list of commonly heard affixes is a short one when it includes affixes whose dependence range is generally within a phrase span — i.e. within a bracket. Two affixes in this first list appear again in an even shorter second list which includes a few other affixes whose dependence range is wider-extending to whole clause or even to whole sentence.

The pluralizing suffix appears not only after pronoun, as [I-pluralizer] meaning we (uo^3 men^0), but also after some nouns in a sociolinguistic set whose generic is man as a human being, including status and role terms, as professor, student, worker, comrade in M-m sequence and in compound sequence, as (M - M) - m. Since nouns are non-committal as to number, the pluralizer suffix at least adds explicitness of number (in a collective sense). Thus, [ren^2] person or persons but [ren^2 men^0] womenfolk, people. Quantification by prior numeral is incompatible with pluralizing by suffix: [three classifier person] meaning three people (san^1 ke^0 ren^2), but never

*

*[three classifier person pluralizer]. On the other hand, pronouns do mark singular person, and hence are really changed in number by the pluralizer suffix. When the pronoun referent is to human beings, number is specified; number is non-committal, however, in pronominal reference to plants, to inanimate objects, or to some animals. Thus, ta^{1} means it or they in reference to non-domesticated animals, plants, and things; but in reference to humans and domesticated animals, its meaning is restricted to third person singular of either sex, as he ~ she. And the referent of the M-m sequence is restricted to human (or domesticated animal) third person plural, as they (ta^{1} men^{0}).

The pluralizer suffix has least to do with syntax, since its dependence range is restricted to the pronoun or noun preceding it, in contrast to the subordinating affix, te^{0}, which appears in many environments. Thus, it appears in [M - m M] sequence where the M - m is modifier to the following M, either as possessive modifier or as adjective modifier. Thus, with pronoun, [he-of book] means his book ~ books (t'a^{1} te^{0} šu^{1}); so also with noun, [horse-of foot] means horse's foot ~ feet (ma^{3} te^{0} čiau^{3}). But as adjective modifier before noun, [gold-of money] means gold coin (čin^{1} te^{0} č'ien^{2}) in contrast to the noun compound where gold enters the compound meaning money; hence (money money) rather than [(gold money)] which means wealth, treasure, capital in the form of money (čin^{1} čien^{2}). The same suffix appears in [(M ± M) - m] sequence where the suffixed reduplication is adverbial modifier of the following verb phrase, which happens to be immediately following in sentence.

diately following in sentence (18),

He runs quickly.

(18) t'a^1 k'uai^4 k'uai^0 te^0 p'au^3

he (quick ± quick) -ly run

$[M]_a$ $[(M \pm M)$ $-m]_b$ $[M]_c$

Between the 'adverbial modifier' and the 'following verb phrase', there is an intervening bracket in sentence (19) which encloses a localizing phrase,

He is running quickly on the street.

(19) ta^1 k'uai^4 k'uai^0 te^0 cai^4 ma^3 lu^4 šaŋ^0 p'au^3

he (quick ± quick) -ly is at (Macadam road) top run

$[M]_a$ $[(M \pm M)$ $-m]_b$ $[\frac{m}{M}$ $(M$ $M)$ $\frac{m}{M}]_c$ $[M]_d$

The brackets may be reordered from a b c d as given in sentence (19) to a c b d, without change of message. The same morphemes that appeared in sentence (18) appear again, but in a different order in sentence (20) which has two messages, depending on whether t'a^1 pau^3 te^0 is taken as topic before the comment k'uai^4; or whether ta^1 is taken as topic, and the rest as comment:

His running is fast (i.e. As for him, running is fast);

He can run fast.

(20) t'a^1 p'au^3 te^0 k'uai^4

In the first message He is fast, with modifier in manner of running before fast. In the second message of sentence (20), the suffix te^0 means can, in the sense of potential ability to perform (as the preceding verb specifies) but in the manner specified by a following complement (*fast*). The subject is omitted in sentence (21), which begins with the object:

This kind can be eaten (is edible).

(21)	če^{4}	ke^{0}	č'ɨ^{1}	te^{0}
	this	*classifier*	*eat*	*can*
	$[{m \atop M}$	$m]_a$	$[M\ -$	$m]_b$

Here bracket$_b$ is a verb phrase comment which does not include an embedded actor; while bracket$_a$ is a noun phrase topic functioning as object rather than as subject, which as already mentioned is not specified; hence the object phrase precedes the verb phrase. This is in one respect parallel to sentence (22) which also shows noun in topic before verb in comment, but with verb followed by final particle suffix -le^{0}, and with two possible messages, depending on whether the *chicken* in the topic is taken as prepared food, in which case it functions as the object of the verb; or as a living organism, in which case it functions as subject:

As for chicken (pause), *not eating any more*; or

The chickens are not eating any more.

(22)	či^{1}	pu^{4}	č'ɨ^{1}	le^{0}
	chicken	*not*	*eat*	*any more*
	$[M]_a$	$[m\ -$	$M]_b$	m

As a final particle suffix, -le^{0} is glossed *any more* in the sense of *that's the situation*. In the first message, the topic functions as the object of the following verb phrase comment, with actor not specified, in the order of object-verb which is understood as such only because there is a pause between object as topic and verb as comment; if uo^3 for actor *I* were inserted, the expanded sentence (22) would unambiguously reflect the object-subject-verb order, *chicken I am not eating*, analogous to *money I have*. In the second message, the topic functions as the subject of the verb phrase comment. In both messages the dependence range of the -le^0 as final particle suffix extends over both topic and comment, and hence over the entire sentence. There are some other final particle suffixes which appear in sentence final or after topic, or after coordinate comment brackets, with the same wide dependence range — ma^0 *interrogative*, a^0 *exclamation*, pa^0 *mild imperative*, *tentative* or *suppositional*, and a few others.

So far, -le^0 has been shown out of bracket, as a final particle; but it also appears within bracket when it is glossed as *perfective*. So far, also, -te^0 has been shown within bracket when it is glossed as *can*, *potential ability* in some sequences, and as *subordinator* for modifiers in sequences in which modified follows; but -te^0 also appears out of bracket, as a final particle suffix, when it is glossed as *that's it*, as in

The guest is just apt to come (that's the situation).

(23)	k'e^4	ren^0	huei^4	lai^2	te^0
	(*guest*	*person*)	*can*	*come*	*that's it*
	[(M	M)]	[$\frac{m}{M}$	$\frac{M}{m}$]	m

There is no ambiguity in glossing the suffix -le^0 as *perfective* when it occurs within bracket while glossing it as *oh look*, *obviously*, *as you can see*, *don't* you see, when it occurs as a final particle suffix, out of bracket; both occur in, He has eaten noodles (as you can see).

(24) t'a^1 č'ɨ^1 le^0 mien^4 le^0
he eat perfective noodles
$[M]_a$ $[M - m]_b$ $[M]_c$ m

The brackets in sentence (24) may be reordered from a b c to c a b for Noodles he has eaten, but the out of bracket m representing the final le^0 does not appear in this reordering; in fact, final le^0 as particle out of bracket never follows suffix le^0 in juxtaposition or, conversely stated, suffix le^0 never precedes out of bracket final le^0. There is no ambiguity in the message of sentence (24), or in any sentence including transitive verb with object specified or intransitive verb with certain kinds of complements specified; in the sequence M - m at the end of a sentence, with M representing verb and m representing le^0, the le^0 is perfective. But in sentences including transitive verb without complement, in the same M - m sequence at end of sentence, the m may represent either le^0 as perfective within bracket or the le^0 as final particle suffix, out of bracket. Compare the two messages,

The guest has come; or (Oh look), a guest is coming.

(25) k'e^4 ren^0 lai^2 le^0
(guest person) come
[(M M)] [M - m] = has come (perfective)
[M] m = is coming (new situation).

These two suffixes, te^0 and le^0, belong in both the first list of affixes (with dependence range within the phrase enclosed by brackets),, and in the second list of affixes (out of bracket, with dependence range extending over the whole sentence). Other affixes in the second list appear exclusively out of bracket (see above). Other affixes in the first list appear exclusively within brackets, as the suffix če which marks durative, progressive, and the suffix men^0 which marks plural of pronouns and human nouns, as shown above. Prefixes (minor morphemes) in the first list include hen^3 very, and pu not, when in self-sufficient sentences, and a few others which occur in such sequences as m-M, m-m-M, and m-m-m-M, where m represents a prefix, while M almost always represents a verb. But hen^3 very appears as a major morpheme in [hau^3 te^0 hen^4] good to an extreme. Compare different comments after the topic [t'a^1] he, as [not come] (pu^0 lai^2); and as [also come] (ie^3 lai^2), and as [future again come] (cai^4 lai^2), on the one hand; and on the other hand, the same three prefixes in one comment: He won't ever come again (that's the situation).

(26)	t'a^1	cai^4	ie^0	pu^0	lai^2	le^0
	he	future again	also	not	come	that's it
	[M]$_a$	[m-	m-	m-	M]$_b$	m

The first m- in bracket b, representing cai^4 future again would not appear in a sentence like this without final particle suffix, out of bracket. And so in general, when not in dependent clause, čai^4 co-occurs with final particle suffix — as le^0 in sentence (26), and as pa^0 in (27):

Let's chat a little more.

(27) cai^4 t'an^2 t'an^0 pa^0

future again (chat chat) advisative

$[m- \quad (M \pm M)] \quad m$

The following sentences touch on an interesting distinction between 'think verbs' (hope, expect, believe), which may have clauses — precedable by pause — as objects, and 'link verbs' the objects of which have pivotal function.

I don't think he'll come again.

(28) uo^3 pu^1 ṣ̌iaŋ^3 t'a^1 cai^4 lai^2

I not think he future again come

$[M]_a \Big[[m- \quad M]_x \quad [M]_y \quad [m- \quad M]_z\Big]_b$

Bracket a represents the sentence topic, bracket b the sentence comment. But within bracket b are brackets x, a transitive verb phrase relating the sentence topic to the object clause, brackets y and z. Within this clause, bracket y represents an embedded clause topic, and bracket z an embedded clause comment.

Compare sentence (28) with a 'think verb' and (28′) with a 'link verb'.

I don't want him to come again.

(28′) uo^3 pu^1 iau^4 t'a^1 cai^4 lai^2

I not want he future again come

$[M]_a \Big[[m - M]_x \quad [M]_y \quad [m- \quad M]_z\Big]_b$

Here ta^1 functions both as object of the 'link verb' want and as subject of come; hence, brackets y and z no longer form a clause as in (28).

A pivotal construction after a link verb also appears in the following sentence of command, which is structurally a sentence comment without preceding sentence topic:

Ask him to come again.

(29)	č'iŋ^{3}	t'a^{1}	cai^{4}	lai^{2}
	invite	he	future again	come
	$\Big[[M]_x$	$[M]_y$	$[m$	$m]_z\Big]_a$

Bracket a represents the sentence comment; within bracket a is bracket x, a transitive verb whose object is represented by bracket y, which also serves as subject of z.

He did right in beating that dog.

(30)	t'a^{1}	ta^{2}	nei^{4}	čɨ^{0}	kou^{3}	šɨ^{4}	iŋ^{1}	kai^{1}	te^{0}
	he	beat	that	classifier	dog	is	ought	ought	that
	$\Big[[M]_x$	$[M]_y$	$[\genfrac{}{}{0pt}{}{m}{M}$	m	$M]_z\Big]_a$	$\Big[[M]$	$[(M-$	$M)$	$m]\Big]_b$

Sentence topic, as a whole, functions as a clause subject. It includes embedded clause topic, x, and embedded clause comment, y and z. Though embedded in this sentences , the clause is not a dependent one because it is possible to say [he] [beat] [that classifier dog] independently. Sentence comment b , however, appears as cited only after a clause-included sentence topic — that is, after a sentence topic which marks action, as dog beating in sentence (30). If, instead, the sentence topic included a noun or noun phrase rather than a clause with verb, comment b would be altered from the selection cited ([(ought ought) that]) to the following: tuei^{4} te^{0} [correct that], with the translation altered from an ethical judgement, He did (or does) right in beating that dog — the dog deserved a beating — to a statement of approval, in the sense that the actor was (or is) judged to be quite correct.

At the sentence level, variability in length depends on whether a sentence includes topic and comment, or comment alone; and on whether a clause is included in either topic, or comment, or both. At a lower-than sentence level, phrase brackets are said to be short when they enclose one or two major morphemes (or parentheses for compound or reduplication); phrase brackets are said to be long when they enclose more than two major morphemes.

Where a sentence is of the interrogative kind in which the question is marked by paired alternate choice, the bracket span or phrase is generally short. The subscript letters for the brackets specifying alternate choice in the following sentences are b for the first choice, and b-prime (b´) for the second choice.

Are you going?

(31) ni^{3} č'ü^{4} pu^{0} č'ü^{4}

you go not go

[M]$_{a}$ [$\frac{M}{m}$]$_{b}$ [m $\frac{M}{m}$]$_{b'}$

The minimum answer to this question is either bracket b (for the affirmative), or bracket b´ (for the negative) — either go, or not go. Compare alternate choices without negative, as in

Does he want tea or coffee?

(32) t'a^1 iau^4 č'a^2 hai^2 šɨ^0 iau^4 k'a^{1}fei^1
he want tea still be want coffee
[M]$_a$ [$\frac{M}{m}$ M]$_b$ -[m- M]- [$\frac{M}{m}$ M]$_{b'}$

The minimum answer for this question is either bracket b for the first choice, or bracket b′ for the second choice — either want tea or want coffee. Where the object of the verb is included in the alternate choice brackets, a bilaterally dependent phrase bracket meaning or, -[hai^2 šɨ^0]- or -[šɨ^0]-, often intervenes between the brackets or (rarely) precedes the alternate choice brackets. The alternate choice brackets may also be juxtaposed and the bilaterally dependent phrase hai^2 šɨ^0 may be omitted — without changing the message. So also in,

Do you like to eat rice or eat noodles?

(33) ni^2 ši^3 huan^1 č'ɨ^1 fan^4 hai^2 sɨ^0 č'ɨ^1 mien^4
you (joy -delight) eat rice still be eat noodles
[M]$_a$ [(M - M)]$_b$ [M M]$_c$ -[m - M]- [M M]$_{c'}$

The most likely answer to this question is either brackets b and c for the first choice, or brackets b and c′ for the second choice — either like eat rice or like eat noodles.

Sentence (31) is briefer than the following sentence (34), in which a directional phrase is in a dependence relationship to each of the preceding alternate choice brackets, and is therefore enclosed in a separate bracket, discontinuous with the first choice bracket b. This directional phrase cannot be included exclusively within the second choice bracket b′ because its dependence range extends equally to both brackets, as the minimum answers show.

Is he jumping down?

(34) t'a^1 t'iau^4 pu^1 t'iau^4 sia^4 č'ü^4
he jump not jump down away
$[M]_a$ $[M]-_b$ $[m-$ $M]-_{b'}$ $-[\frac{m}{M}$ $-\frac{m}{M}]_c$

The minimum answer to this question is either brackets b and c (for the affirmative), or brackets b′ and c (for the negative) — either jump down away, or not jump down away.

Most phrase brackets in the alternate choice sentences, (32), (33), (34), include two morphemes. When a sentence is of the kind in which the question is marked by an interrogative demonstrative, the bracket span or phrase is often shorter.

Whom did he beat?

(35) t'a^1 ta^3 šei^2
he beat whom
$[M]_a$ $[M]_b$ $[M]_c$

In a possible answer to this question, brackets a and b remain unchanged while, for bracket c, a pronoun (uo^3 me) may replace the interrogative demonstrative — he beat me.

Who beat him?

(36) šei^2 ta^3 t'a^1
who beat him
$[M]_a$ $[M]_b$ $[M]_c$

In a possible answer to this question, brackets b and c remain unchanged while, for bracket a, a pronoun (uo^3 I) may replace the interrogative demonstrative — I beat him. And so also for other interrogative demonstratives (as na^3 li^0 where, šem^2 me^0 what, na^3 which): the order of brackets in the question sentence remains parallel to that of the answer sentence which differs from the question only in substituting a noun or pronoun for the interrogative demonstrative. Like other demonstratives, the interrogative demonstrative also appears as a prior member before classifier relevant to a following noun (see above).

However, where a sentence is of the interrogative kind in which the question is marked not by paired alternative choice, and not by interrogative demonstrative, the bracket span or phrase either is, or may be, long — that is, longer than one or two major morphemes or their compounded or reduplicated equivalents. The same four morphemes appear below in sentence (37), bracket b (prior to the interrogative final particle suffix, ma^0 out of bracket), and in sentence (38), bracket b, (where the interrogative is marked by giving an alternative choice in bracket c — not, however, a coordinately paired choice).

Has he jumped down?

(37)	t'a^1	t'iau^4	šia^4	čü^4	le^0	ma^0
	he	jump	down	away	perfective	interrogative
	$[M]_a$	$[M$	$\frac{m}{M}$	$\frac{m}{M}$	$m]_b$	m

(38)	t'a^1	t'iau^4	šia^4	čü^4	le^0	mei^0	iou^0
	he	jump	down	away	perfective	not	have
	[M]$_a$	[M	$\frac{m}{M}$	$\frac{m}{M}$	m]$_b$	[$\frac{m}{M}$	M]$_c$

Though the question is the same, whether expressed as sentence (37) or (38), the difference between these two sentences is appreciable in terms of the different minimum affirmative-negative answer or response that would follow. For sentence (37), it would be affirmative it's that (šɨ^4 te^0), or negative has not (mei^2 iou^0). For sentence (38), the minimum answer would be either bracket b affirmatively, or bracket c negatively — either jump down away perfective, or has not.

Many sentences show two successive brackets enclosing noun and verb phrases — the topic (as subject) and the comment (as predicate). But the topic cannot always be equated to subject in function, nor the comment to the predicate in function. When the topic functions as subject and the comment as predicate, they do so as one special function out of the various functions of topic and comment.

At a lower level, bracket phrases are often short, especially so in alternate choice paired questions and their minimum answers, and in imperative sentences. When the brackets are long, the order of morphemes within the bracket is fixed in reference to a nucleus verb or noun (which may be a nucleus major morpheme, M, or nucleus parentheses enclosing reduplicated of compounded M's). Though both short and long brackets, as wholes, are often reorderable or permutable within a given topic or comment, the order of morphemes within any given bracket is no more reorderable than are affixes in languages with a 'word' level. Indeed, the bracket as a whole is analogous to a 'word' in this sense. But in another sense, it is analogous to a phrase. The nucleus noun or verb in the bracket may be flanked not only by affix-like minor morphemes (m, in our notation), but also by morphemes which function as minor morphemes, $(\frac{m}{M})$, when flanking a nucleus major morpheme. A hundred or so classifiers may appear in the same bracket before noun nucleus just as a smaller inventory of localizers may follow the noun nucleus, again in the same bracket, as has already been shown. But the very same morpheme that functions as localizer $(\frac{m}{M})$ after noun nucleus, or as classifier $(\frac{m}{M})$ before noun nucleus, may also function as the noun or verb nucleus $(\frac{M}{m})$ in another bracket, as is most clearly shown in sentence (4a), where feŋ^{4}, as verb nucleus, means *to seal* $(\frac{M}{m})$ in one bracket, while feŋ^{0} means *classifier relevant to envelope or small package* $(\frac{m}{M})$ in another bracket.

In general, the relationship of prior morphemes to the nucleus of the bracket is predominantly that of modifier-modified — both when the modifier (or modifier sequence) is adverbial before a verb nucleus, and when it is adjectival before a noun nucleus in the bracket. Examples of the latter are *big horse* [ta^{4} ma^{3}], *small house* [čia^{1} (faŋ^{2} cɨ^{0})], and *good man* [hau^{3} ren^{2}]. The adjective noun sequence is in effect an expanded noun phrase (as in sentence 61).

There is no predominant, general relationship of following morphemes to the preceding noun or verb nucleus, though one of the most clear-cut relationships is that of localizer after the noun (as house in [čia^1 li^0] for in the house), and of directional after the verb — but many other kinds of morphemes may also follow in the same bracket, including suffixes (see above).

The term 'verb complement' is used for all morphemes of the $\frac{m}{M}$ type that follow the phrase nucleus in the same bracket. When a single verb complement or a pair of such complements specifies direction of an active verb, it is termed a 'directional'. The same morpheme is glossed differently when it functions as a directional in one bracket [:··$\frac{m}{M}$] than when it functions as a nucleus verb in another bracket [$\frac{M}{m}$ ···] (in most instances with change of tone). Thus [··· *here, hitherward*] as directional, but [*come* ···] as nucleus verb (lai^2); [*there, away*] as directional, but [*go* ···] as nucleus verb (č'ü^4); [··· *down*] as directional, but [*descend* ···] as verb (šia); [··· *up*] as directional, but [*ascend* ···] as verb (šaŋ); and [··· *across*] as directional, but [*pass by, pass across* ···] as verb (kuo). The directional may be followed by a suffix (as the perfective suffix in the following examples). Compare *he ran* (t'a^1 p'au^3 le^0) with

He ran here.

(39)	t'a^1	p'au^3	lai^0	le^0
	he	*run*	*hither*	*perfective*
	[M]	[M'	$\frac{m}{M}$	m]

Sequences of two directionals after active verbs also occur marking both *up* (or *down* or *across*), and either *hitherward* or *away* (from speaker).. Compare *he passed by* — t'a^1 kuo le^0 [M] [$\frac{M}{m}$ m] — with

He ran across that way.

(40)	t'a^1	p'au^3	kuo	č'ü^0	le^0
	he	*run*	*across*	*away*	*perfective*
	[M]	[M	$\frac{m}{M}$	$\frac{m}{M}$	m]

Compare he has eaten it (t'a^{1} č'ɨ^{1} le^{0}) with sentence (41) which might be uttered when formerly visible food had been all eaten, so that no food remained to be seen.

He has eaten it up completely.

(41)	t'a^{1}	č'ɨ^{1}	šia^{0}	č'ü^{0}	le^{0}
	he	eat	down	away	perfective
	[M]	[M	$\frac{m}{M}$	$\frac{m}{M}$	m]

Sentences (42), (43), (44), which follow, exemplify the triple referents or meanings of two morphemes which are members of a single compound, (č'i lai) — one referent when the compound functions as the verb nucleus (42), and two different referents when the compound functions as a verb complement, depending on whether the verb complement follows an active nucleus verb (43), or a stative nucleus verb (44). It means up after active verb, but begin, start after stative verb, when it is called a 'resultative complement' rather than a 'directional complement'. As a directional complement, the compound is glossed literally as (upward hitherward). As a resultative complement, it is glossed (inceptive resultative). As the verb nucleus itself, the same morphemes in the compound are glossed (rise come). The examples which follow, let us repeat, are instances of a single compound with three different meanings in three different environments; this single compound is not to be confused with sequences of two directionals which are not members of a compound (40, 41).

He arose (got up from bed).

(42) t'a^1 č'i^3 lai^2 le^0

he (rise come) perfective

[M] [($\frac{M}{m}$ $\frac{M}{m}$) m]

He has stood up

(43) t'a^1 čan^4 č'i^0 lai^0 le^0

he stand (upward hitherward) perfective

[M] [M ($\frac{m}{M}$ $\frac{m}{M}$) m]

He has started to laugh

(44) t'a^1 šiau^4 č'i^0 lai^0 le^0

he laugh (inceptive resultative) perfective

[M] [M ($\frac{m}{M}$ $\frac{m}{M}$) m]

When the subject of a sentence is included in the topic, the observable movement of the subject is minimized by stative verb comment (marking to laugh, to sing, to sleep, etc.) but not by active verb comment (marking to run, to eat, to swallow, to walk, etc). Some stative verbs are incompatible with topic functioning as subject. And they occur generally in comment sentences, without any prior topic at all. The compound verb to rain (šia^4 ü^3) is an example of such a stative verb. Accordingly, the compound verb complement (č'i^0 lai^0) is glossed (inceptive resultative) -- rather than * (upward hitherward)-- when combined with the compound verb (šia^4 ü^3), meaning it started to rain (rather than *it rained up here). In general, the happenings of natural phenomena are specified by stative verbs which are incompatible with prior topic functioning as subject.

Some verb compounds appear not only with the constituent morphemes in a juxtaposed sequence, as (šia^4 ü^3) it's raining, but also in a discontinuous sequence, with the first member of the compound (šia^4) — separated from the second member of the compound — (ü^3) by an interspersed or infixed morpheme (as perfective le^0); le^0 could also be added as the final morpheme of sentences (45) and (46), just as it could have been inserted after č'ɨ^1 in sentence (41).

It has rained.

(45) šia^4 le^0 ü^3

descends perfective rain

$[(\frac{M}{m})$ - m -(M)]

In general, two member verb compounds of the kind glossed (sleep-nap) and (sing song) are separable in the same way. Compare both [he] (sleep-nap) for he sleeps (t'a^1 šuei^4 čiau^4), and [he] [(sing song)] for he sings (t'a^1 č'aŋ^4 ke^1), with sentence (46); the perfective le^0 may also be infixed between sing (čaŋ^4)- and song-(ke^1), as another example of discontinuous compound.

He slept.

(46) t'a^1 šuei^4 le^0 čiau^4

he sleep perfective nap

[M] [(M)- m -(M)]

The compound verb complement, (č'i^0 lai^0), appears as a continuous compound after stative verb (38); it is also separable, but in a new way. What intervenes between (č'i^0) = and = (lai^0) is the second member of a compound verb. Then the first member of the compound verb precedes the first member of the compound verb complement in an interlocked or interwined sequence. [( )- (č'i^0) = -( ) = (lai^0)]:

It's beginning to rain — that's the situation.

(47)	šia^4	č'i^0	ü^3	lai^0	le^0
	descend	inceptive	rain	resultative	that's it
	[(M)-	($\frac{m}{M}$) =	-(M)	=($\frac{m}{M}$)]	m

In this sentence, the final particle suffix, le^0, is out of bracket and yet, though merely background, a necessary 'ground' for the bracket 'figures' in the foreground. (see above). Within the bracket, the compound nucleus verb (šia^4 ü^3) combines with the compound verb complement (č'i^0 lai^0) in a discontinuously intertwined fashion — as if two double-pronged gestalt figures, ┌──┐ and └──┘, were integrated in one interlocking figure, ┌└─┐┘.

Nucleus verb compounds, like that for (descend rain) above, combine with verb complement compounds only by integration — by ┌└─┐┘ interlocking. Nucleus verb compounds like those for (sing song) and (sleep nap) may be reduced to their first members before the compound verb complement, as in sentence (48), or they may remain compounds and integrate with the compound verb complement by ┌└─┐┘ interlocking, as in sentence (49).

He started to sing.

(48)	t'a^1	č'an^4	č'i^0	lai^0	le^0
	he	sing	inceptive	resultative	perfective
	[M]	[M	$(\frac{m}{M}$	$\frac{m}{M})$	m]

(49)	t'a^1	č'aŋ^4	č'i^0	ke^1	lai^0	le^0
	he	sing	inceptive	song	resultative	perfective
	[M]	[(M) -	$(\frac{m}{M})$ =	-(M)	= $(\frac{m}{M})$	m]

Some compounds are potentially separably, in the sense that certain other morphemes in the comment may intervene between the members of the compound thereby making the compound a discontinuous one (see above). Other compounds are characteristically discontinuous — in other words, discontinuous whenever possible. Thus, the compound verb complement for out hitherward occurs as a continuous compound only in short comments, as in the following sentences: (50) and (51). When the comment is longer, by virtue of including the object of a transitive verb, this object appears either as a short inserted bracket (sentence (52)), or as a long inserted bracket (sentence (53)), between the first member of the directional compound and the second member. The latter is not really a separate phrase, but rather a discontinuous appendage of the first phrase in bracket b. In the following sentences, the comment follows either topic as subject — (50), (52), (53); or topic as object — (51).

He walked out (toward speaker).

(50)	ta^{1}	cou^{3}	č'u^{0}	lai^{0}
	he	walk	out	hitherward
	$[M]$	$[M$	$(\frac{m}{M}$	$\frac{m}{M})]$

Put the money out here (with pause after č'ien^{2}).

(51)	č'ien^{2}	na^{2}	č'u^{0}	lai^{0}
	money	take	out	hitherward
	$[M]$	$[M$	$(\frac{m}{M}$	$\frac{m}{M})\ ]$

He put the money out (toward speaker).

(52)	t'a^{1}	na^{2}	č'u^{0}	č'ien^{2}	lai^{0}
	he	take	out	money	hitherward
	$[M]_{a}$	$[M$	$(\frac{m}{M})-$	$[M]_{c}$	$-(\frac{m}{M})]_{b}$

He took out a book.

(53)	t'a^{1}	na^{2}	č'u^{0}	i^{0}	pen^{2}	šu^{1}	lai^{0}
	he	take	out	one	classifier	book	hitherward
	$[M]_{a}$	$[M$	$(\frac{m}{M})-$	$[\frac{m}{M}$	$\frac{m}{M}$	$M]_{c}$	$-(\frac{m}{M})]_{b}$

Compare sentence (52), above, and sentences like (54) and (55) where—though the subject is not specified — the verb is still active rather than passive.

As for the rice (pause), it has been eaten up.

(54)	fan^{4}	č'ɨ^{1}	uan^{2}	le^{0}
	rice	eat	finish	perfective
	$[M]_{a}$	$[M$	$\frac{m}{M}$	$m\ ]_{b}$

As for the bowl (pause), it got broken.

(55)	uan^{3}	ta^{3}	p'uo^{4}	le^{0}
	bowl	beat	broken	perfective
	$[M]_{a}$	$[M$	$\frac{m}{M}$	$m\,]_{b}$

In the free translation, it is added for smoothness. The actor is simply not specified in sentences (54) and (55)—and these are clearly not passive sentences, but sentences in which topic, a, functions as object and comment, b, includes a transitive nucleus verb, followed by verb complement and perfective suffix.

Sentence (56) shows a passive construction in 'acquired Mandarin' with pei^{4} before a transitive verb nucleus. Compare a possible answer to the question By whom was he beaten? — pei^{4} t'a^{1} by him. An active, rather than passive, causative construction is expressed in sentence, (57) with pa^{0} before a transitive verb nucleus.

The dog has been beaten by him.

(56)	kou^{3}	pai^{4}	t'a^{1}	ta^{3}	le^{0}
	dog	by, be caused	he	beat	perfective
	$[M]$	$[\frac{m}{M}$	M	M	$m\,]$

(As a major morpheme in compound, pei^{4} is glossed to be covered, to cover.)

He has broken the bowl.

(57)	t'a^{1}	pa^{0}	uan^{2}	ta^{3}	puo^{4}	le^{0}
	he	by handling	bowl	beat	broken	perfective
	$[M]$	$[\frac{m}{M}$	$M\,]$	$[\,M$	$\frac{m}{M}$	$m\,]$

These two sentences differ in the function of the topic — with topic as subject (57); and with topic as object of the verb ta^{3} beat (56).

When the nucleus verb is preceded by one verb in the domain of travel (as walk, run, go, come), the relation of the prior verb to the nucleus verb is simply purposive (in order to).

I'm going in order to see him.

(58)	uo^{3}	č'ü^{0}	k'an^{4}	t'a^{1}
	I	go	see	him
	[M]	[M	M]	[M]

Compare huei^{4} can, know how to, followed by verb, as in I know how to sing [uo^{3}] [huei^{4} (č'aŋ^{4} ke^{1})] .

The comment may consist of a series of brackets, each with a nucleus verb, which may be flanked by another verb dependent on the nucleus verb (and/or by prior adverbial, and/or by following verb complement). The different brackets in a successive series of brackets are coordinately related, and hence the brackets in itemization series are maximally reorderable.

Every day he reads and eats and sleeps.

(59)	t'a^{1}	mei^{2}	t'ien^{0}	tu^{2}	šu^{1}	č'ɨ^{1}	fan^{4}	šuei^{4}	čiau^{4}
	he	every	day	read	book	eat	rice	sleep	nap
	[M]$_{a}$	[M	M]$_{b}$	[(M	M)]$_{c}$	[(M	M)]$_{d}$	[(M - M)]$_{e}$	

Sentence (59) is reorderable to Every day he reads and sleeps and eats; and to Every day he sleeps and reads and eats; and to Every day he eats and sleeps and reads; and so on.

In contrast to the brackets in successive series, above, brackets b and c in sentence (60), below, are not coordinate because the nucleus verb in each is differently flanked. In c, the nucleus verb is flanked by prior purposive verb go in order to; in b, the nucleus verb is flanked by following verb complement and its object. The free translation telescopes run and its complement, reaching, which can only have a place as object, and the purposive go into two words, went to.

I went to his house to see you.

(60)	uo^{2}	p'au^{3}	tau^{0}	t'a^{1}	čia^{1}	čü^{4}	k'an^{4}	ni^{3}
	I	run	reaching	his	house	go	see	you
	$[M]_a$	$[M$	$\frac{m}{M}$	M	$M]_b$	$[M$	$M]_c$	$[M]_d$

A short bracket encloses one or two morphemes or parentheses (compounds and/or reduplications). A long bracket encloses three or more such units. A long bracket with verb nucleus is naturally restricted to about three or four such units, because longer stretches than this would be segregated into successive brackets, each with its own verb nucleus, as in sentences (59) and (60), above.

Brackets with noun nucleus may be somewhat longer, in their maximum, than brackets with verb nucleus. This is because it is possible to expand a demonstrative (or numeral)-classifier- noun bracket, by the insertion of a single adjective (or an adjective sequence) between the classifier and the noun nucleus— as big between two-classifier and dog for two big dogs (liaŋ^{3} čɨ^{0} ta^{4} kou^{3}). When a longer adjective sequence is inserted, all morphemes preceding are in a modifier relationship to the modified nucleus noun (ren^{2} man, in the following sentence). As many as a half dozen morphemes, more

or less, may precede the nucleus noun — five single morphemes and one compound in

Three men wearing black clothes are dining.

(61)	**san^{1}**	**ke^{0}**	**č'uan^{1}**	**hei^{1}**	**i^{1}**	**fu^{0}**	**te^{0}**
	three	**classifier**	**wear**	**black**	**(clothes**	**clothes)**	**'s**
	[M	**m**	**M**	**M**	**(M-**	**M)**	**m**

ren^{2}	**cai^{4}**	**č'ɨ^{1}**	**fan^{4}**
man	**at**	**eat**	**rice**
M]	**[$\frac{m}{M}$**	**(M**	**M)]**

The preceding sentences lend themselves to a summary of text frequency, with the count based on the 61 sentences as cited above. Thus, the number of occurences of morphemes in the 61 sentences is 333, where a given morpheme (as ta^{1} *he*) is counted separately every time it occurs, rather than once for each morpheme (as might be done in an inventory count).

Each morpheme in the 61 sentences falls into one of three classes (M, m, or $\frac{M}{m}$ alternating with $\frac{m}{M}$).

Most of the morphemes in these sentences - 58 percent - are always major morphemes (M's); and here the text frequency approximates an inventory count. (It seems likely that more than half of the total lexical resources of 'acquired Mandarin' would function invariably as major morphemes.)

Surprisingly high in text frequency are the affix-like minor morphemes (m's). Almost one fifth of the 333 morphemes function as such; 17 percent are always minor morphemes. But in inventory count, there are only about a dozen minor morphemes, which occur again and again in the different sentences - for example le^{o} appears 23 times, te^{o} appears 12 times, and so on to ie^{o}, pa^{o}, and ma^{o} which occur once each in text frequency. In other texts than our sample of 61 sentences, the latter might show a much higher text frequency. In general something less than one quarter of the morphemes in any text might function invariably as minor morphemes (m's); and as noted above, something more than one half function invariably as major morphemes (M's).

That leaves one quarter of the 333 morphemes which function as minor morphemes ($\frac{m}{M}$) in some environments and as major morphemes ($\frac{M}{m}$) in others. Exactly 82 out of the 333 - 25 percent of all morphemes in text frequency - fall in this alternable $\frac{m}{M} \sim \frac{M}{m}$ class. Accordingly, about half of the 333 morphemes (the minor morphemes, m, and the alternable $\frac{m}{M} \sim \frac{M}{m}$ classes), are as relevant to 'morphology' as to 'syntax'.

At the next higher level of grammar, compounds and/or reduplications are enclosed in parentheses. Such parentheses are found in more than half of the sentences - in 33 of the 61 sentences. In some sentences, however, a single morpheme or two is flanked by compounds and reduplications; thus, the sequence for sentence 13 is "(reduplicated classifier)-(compound noun)-single morpheme glossed all-(compound predicate adjective)".

One member of a compound is frequently bound - i.e. found only in compounds but both members of a compound may be free (occuring either in compounds or as single morphemes). Less commonly, both members of a compound are bound morphemes. Repeated members of a reduplication are more often free than bound. Of the 33 sentences showing reduplication and/or compounding, scarcely more than half a dozen parentheses enclose a bound morpheme. Here text frequency shows that less than 20 percent of the compounds include a bound member. What would an inventory count of a dictionary show?

The parenthesis span, which includes two or more morphemes, whether bound of free, functions as one 'part of speech' just as each single major morpheme out of parenthesis functions as one 'part of speech'. In the functional view then, the parenthesis span and the single major morpheme are at the same word-like level (since each can be classified as 'noun' or 'verb' or 'adjective' or 'adverb' - and in addition some single major morphemes may be classified as 'pronoun' part of speech). What makes this word-like level difficult to maintain is that at a still higher level in the grammar, the phrase bracket is also word-like in terms of major 'parts of speech' (since the bracket phrases distinguish the same parts of speech that single major morphemes distinguish, including 'pronoun').

At the intrabracket level, relative order of single morphemes and parenthesis spans is very fixed or restricted, not unlike 'morphology' in languages with a 'word' level. It is at the higher interbracket level that contrastive and non-contrastive reorderings of 'syntax' are found, while 'morphology' is a matter of the fixed relative order at the intrabracket level. But 'morphology' and 'syntax' are telescoped into one whenever a single morpheme (M or $\frac{M}{m}$) is the only member of a phrase bracket, as happens in 69 instances out of a total of 225 phrase brackets in the 61 sentences. Well over half of the phrase brackets - altogether 156 - include two or more morphemes. And this means, in short, that most phrases in the 61 sentences involve problems in 'morphology', while all phrase brackets are involved in 'syntax'. In the 156 phrases that include two or more morphemes, a single morpheme or parenthesis (compound and/or reduplication) serves as the nucleus of the phrase bracket, and the other morphemes in the phrase are in a fixed order relative to this nucleus.

In short phrase brackets, a single morpheme precedes or follows the nucleus. About 10% of the 156 phrases with two or more morphemes show minor morphemes (m's) preceding the verb nucleus (sentences 3, 6, 15, 16, 22, 23, 27, 28, 29, 31, 32, 33, 34, 38). But major morpheme (M) is also found to precede, as modifier, the phrase nucleus major morpheme(M) - in a couple of instances (sentences 23, 25). Some 6% of the same 156 phrases show minor morphemes following the phrase nucleus (sentences 1, 2, 3, 4, 7, 18, 21, 24, 25, 30, 42).

Over 80% of the 156 phrases with two or more morphemes are long phrase brackets in which one morpheme precedes while another morpheme follows the nucleus morpheme; or a pair of morphemes precedes (or follows) while a single morpheme follows (or precedes) the phrase nucleus; and more than two morphemes may flank the phrase nucleus (as in sentences 26, 37, 38, 40, 41, 43, 44).

Some phrase brackets in sentences which are uttered only in macro-syntax (as 11, 12, 21) appear without major morpheme, as a sequence of two minor morphemes, including classifier. When two or more major morphemes appear in the same phrase, the nucleus morpheme is that morpheme whose 'part of speech' class coincides with the phrase 'part of speech'; thus, in a noun phrase the modified noun is phrase nucleus after the modifier, and in transitive verb phrase with non-permutable object noun, the verb is the nucleus morpheme before the object noun (as in sentences 17, 32, 33). When two major morphemes in the one phrase are both 'verb' part of speech, the second 'verb' functions as phrase nucleus when the first is in the domain of travel (go or come in order to), while the first is phrase nucleus in the domain of appraisal (to like, to want, to wish, etc., as in sentences 58 and 59).

Final particle suffixes out of bracket (as in sentences 22, 23) are merely background, but may be an indespensible 'ground' for the complex 'figures' in some foreground brackets (as in sentence 47). Most brackets, however, whether topic and comment, or comment alone, are self-sufficient in the sense that the background particle suffix is optional.

Two-thirds of the 61 sentences in the sample have topic functioning as actor before comment in which only one bracket functions as verb phrase (including predicate adjective functioning as verb in sentences 6 to 13, and 20). A half dozen additional sentences have comments which include two brackets, each functioning as verb phrase.

In the remaining third of the 61 sentences, the topic may function as object before comment (with embedded subject), or the topic includes an embedded verb before comment with single verb, or comment appears without prior topic, again either with one or two phrase brackets.

With two verb phrases in the comment, the second may negate the first (as in sentence 31: $[\underline{\text{go}}]_a\ [\underline{\text{not go}}]_b$) in order to elicit an answer (either bracket a or bracket b). Or the second verb bracket (b prime) may give an alternative choice to the first verb phrase bracket (as in sentence 32), in order to elicit an answer (either bracket b or b prime). With three verb phrases in comment (as in sentence 33), the first verb bracket (b) is a transitive verb with the paired verb phrases following serving as alternative goals (c and c prime) to elicit an answer (b and c) or (b and c prime); or (as in sentence 59), the three or more verb phrases may be coordinate in a simple series, without offering interrogative alternatives.

The preceding 61 sentences constitute one possible sample of 'acquired Mandarin' which serves as a profile grammar that is isomorphic with grammars of all Han Chinese (but not quite isomorphic with Wenyen). Some differences among the various Han Chinese languages and dialects are noted below. That is, the profile of our 'acquired Mandarin' grammar, though isomorphic with other Chinese languages, is not identical with them. The differences are found mainly in phonology, but are also found in grammar.

1.5. Aside from differences in phonemic inventories, which are given in a following fascicle, the chief diversity among the Han Chinese languages and dialects is found in differences in the shapes of morphemes. Some of these (as for ten: Mandarin šɨ^2 but Cantonese sàp) though different, are still cognate — that is, derived from a single parental morpheme. Other morphemes are not only different in shape but also different in etymon, and yet mark the same referent; thus, who is marked by one morpheme (šei^2) in Mandarin but in Cantonese by another morpheme, or rather by two morphemes (pín kɔ̄ — literally side one). In the case of the referent for clothing, it is Mandarin which commonly uses two morphemes in a compound (i^1 fu^2), while Cantonese commonly uses one (sá·m clothing). A single but different morpheme marks walk in Mandarin (čou^3) and Cantonese (hà·ŋ).

Differences in structure are often associated with different morphemes that different Han Chinese languages select, as when a given morpheme (or succession of morphemes, as mei^2 iou^3 in Mandarin for not have) entails a structural restriction (avoidance of mei^2 in sentence final) not found in another language, as Cantonese where a single morpheme appears both in sentence final and in other parts of the sentence for not have (m̀ǒu). But attention can be focussed on occasional structural differences by comparing the behavior of glosses in the 61 sentences cited for 'acquired Mandarin', with different behavior of glosses in other Han Chinese languages. By the same device, the general sameness in structure among all modern Chinese languages and dialects can be reaffirmed in specific contexts.

In Peking Mandarin, the first of the 61 sentences is revised by the addition of one morpheme (pa^3), which changes the structure of the sentence without changing the message (I have written this letter in a slap-dash manner).

Compare sentence (1), above, and the Peking Mandarin revision:

(1) uo^3 ma^3 ma^0 hu^3 hu^0 te^0 pa^3 če^4 feŋ^1 šin^4
[*I*]$_a$ [(*horse* ± *horse*) (*tiger* + *tiger*) -*ly*]$_b$ [*take* *this* *classifier* *letter*]$_c$

šie^3 le^0
[*write* *perfective*]$_d$

Two phrase brackets (a and b) are identical in the 'acquired Mardarin' version and in the Peking Mandarin version. Phrase bracket c, immediately above, adds pa^3 *take* to what is the last bracket in the original version. The phrase bracket [*write perfective*] is last in the Peking revision (d), but pre-final (c) in the original version. The revised version alone is acceptable in Peking Mandarin; both versions are equally acceptable to our speaker of 'acquired Mandarin'. In addition to the slight difference as to what is included in brackets and the slight shift in order of [*write* *perfective*], the two versions differ importantly in structure. The phrase brackets of the original version can be reordered from a b c d (as cited) to d a b c; whichever order is followed, an additional bracket phrase (e.g. [*at* *table* *up*], meaning *on* *the* *table*) may be inserted between phrases b and c, with expanded message. In the Peking version, the reordering is from a b c d to a c b d—that is, the adverbial phrase, [(*horse* ± *horse*) (*tiger*+*tiger*) -*ly*], may occur between the two verb phrases (c and d), or precede both of them. But in either order of the Peking structure, it would seem unnatural to expand the sentence by a phrase localizing the writing (as, *on the table*), though the sentence might be expanded by other phrases, as [*yesterday*] between topic and first bracket of comment.

The half dozen bracket phrases of sentence (3) — *He has never played* happily at home with his own children — seem to be equally reorderable (and translatable) in all modern Chinese languages. A possible alternative selection of morphemes for bracket b — from [not ever] (pu^1 c'eŋ^2) to [not have] (mei^2 iou^0) — would render the sentence more closely translatable into Wu and Cantonese.

The structure of sentence (4) holds for Han Chinese in general, but with selection of morpheme for letter showing a different etymon in S. Min, for example, with the classifier for letter remaining as a cognate of feŋ^0. In addition feŋ^1 functions as a verb (to seal), and enters into a compound for envelope — (letter sealer); but for envelope (šin^4 feŋ^1), the compound is preceded by another classifier than feŋ^0.

(5) You give that man a hand shows reduplication of the verb in Mandarin and Wu (help help), but not in N. Min and Cantonese where the phrase would be glossed [help a stroke] or more briefly [help stroke] before bracket c [that classifier man]. The latter structure is also possible in Mandarin, as [you] [help one help]; or, with object, [you] [help one time] [that classifier man] .

(6) He's very thin. The topic [he] is followed by comment consisting of adjectival verb preceded by an intensifier, [very thin] , as a self-sufficient sentence. In Cantonese the morpheme for good serves as the intensifier, as in the sense of good and thin. Though the intensifier (hen^3) is distinguished from good (hau^3) in Mandarin, the latter may also serve as an intensifier: t'a^1 hau^3 huai^4 for He's very bad.

[he] [good bad]

As an occasional intensifier, hau^3 may appear in non-reduplicative sequence, first as modifier (very) and then as an adjectival verb (good). In Wu, the intensifier is marked by a two syllable adverb which is more emphatic than

Mandarin for very (hen^3). And the Wu bisyllabic morpheme may be final, or precede a final particle without intervening verb. In Mandarin, the occasional intensifier (hau^3) — but not the usual one (hen^3) — may function as an adjectival verb immediately after the topic in a response sentence. For example, the answer to "Who's good?" might be [t'a^1] [hau^3] he's good; rather than he's very good (6, above).

(7) He's rather thin (poor thing), with commiseration marked by reduplication — (thin ± thin). In Peking and in other national capital cities (as Nanking), the non-syllabic suffix -r (heard only in capital cities) appears where no suffix is found, as completing the reduplication (thin ± thin-r), or as an alternate selection for the syllabic diminuitive -cɨ. In Cantonese and in Wu, the morphemes (suffixes or enclitics) that appear after reduplicated forms are apt to be different than those that appear after the non-reduplicated form of the morpheme. Compare non-reduplicated form in sentence (6) with reduplicated form in (7). What such reduplication marks in Han Chinese generally is a sense of liveliness, as when one says [he] [fat ± fat]-ly to mean not only that He's fat, but that there is something lively and cute about his being fat like a baby.

(8) Try and eat this to see (how it is), with the first phrase reduplicated [eat ± eat] and enclosed in brackets before the second phrase [look], because object, when specified, appears between the transitive verb and the intransitive. This is a very productive pattern that appears in Han Chinese generally. Wu selection is much like Mandarin for Look and see — literally, look one look look, or more briefly [(look ± look)] [look] for Take a look to see (how it is) — k'an^4 k'an^0 k'an^4. Many reduplicated morphemes function as verb before object and before intransitive, then see how it is (k'an^4), with the object

giving a semantic guide to the point in referent range of the transitive verb (as ta^3 for *beat, use instrument, play*). Thus, *Hit him and see* (what the consequences are — as crying, running away, fighting back): [(*beat* + *beat*)] [*him*] [*look*] . If the object is *telephone*, the translation is *Make a telephone call and see*; if mahjong, it is *Play mahjong and see.*

Sentences including classifiers — sentences (10) to (13) — are found in Han Chinese generally. And classifiers extend beyond Sino-Tibetan to unrelated SE Asian languages. But such sentence profiles with classifiers differ from those encountered in Han Chinese.

(10) *This card is pretty* shows a profile in which demonstrative (*this*, substitutable by a numeral) preceeds *classifier* before noun (*card*, or adjective and noun, as *big card*). For this sentence Cantonese would select another etymon for the adjectival verb (*bright* in the sense of pretty or attractive), without changing the classifier sentence profile.

(11) *This one is pretty* shows a profile in which demonstrative (*this*, or else a numeral) preceeds *classifier* (substitutable by another *classifier* for *This kind is pretty*).

(12) *Every one of them is pretty* shows a profile in which the reduplicated classifier, serving as topic [(*classifier* *classifier*)], is followed by comment [*all pretty*], which specifically echoes *every one* by *all*; this is highly productive, but not obligatory for all sentences.

(13) *All the cards are pretty* shows a profile in which the reduplicated classifier preceeds noun as topic, with morpheme for *all* again included in comment.

(14) *This is my world* shows a profile possible in Mandarin (demonstrative alone as topic without following classifier or noun), which does not appear in

Han Chinese generally -- at least not in Min, Wu or Cantonese -- where demonstratives (for *this*, *that*, *which*, etc.) would always be followed by a classifier, as in the profile shown in sentence (11), above.

(15) *I don't have;* (16) *I don't have money;* (17) *I haven't any money*. In each of these three sentences, a negative morpheme (mei^2) occurs before the morpheme for *have* (with and without object) or before the object (without intervening *have*); but the negative (mei^2) never occurs in sentence final in Mandarin. In Min, Wu and Cantonese a single morpheme functions like the Mandarin *not* (mei^2) before *have* (iou^3); hence the two Mandarin sentences, (16) and (17), might be said to merge into one sentence in these non-Mandarin languages, and this single morpheme (e.g. mǒu in Cantonese) may appear as the last morpheme of a sentence, parallel to *I don't have* (15,, above). Beside such verbal negatives, Han Chinese languages also mark negative by a non-verbal morpheme (pu^1 *no*, *not* in Mandarin). Both the latter and the verbal negative (as auxiliary) appear in sentences before verbs and adjectival verbs. In the sentence [I] [negative go] , the meaning is *I have not gone* with verbal negative, but *I'm not going*, sometimes in the sense of *not willing to go* with non-verbal negative. In the sentence *He negative good*, with adjectival verb (hau^3), the meaning is *He's no good* with non-verbal negative, but *He hasn't recovered* with verbal negative. *He's not bad* (pu^1)vs. *It hasn't spoiled, become bad* (mei^2).

(18) *He runs quickly.* The adjective for *quick* (k'uai^4) is transformed into an adverb by reduplication plus suffix in Han Chinese in general. A single suffix (te^0) follows the (*quick*+ *quick*) in our sample of 'acquired Mandarin', but in Peking Mandarin a non-syllabic suffix (-r) is included with each member of the reduplication.

(19) *He is running quickly on the street*. Between the reduplicated ad-

verbial phrase for quickly (as for 18, above) and the final verbal phrase [run], there is a phrase bracket beginning with a co-verb is at (cai^4) followed by a place expression which includes a compound for street — (horse road) in popular etymology, because ma^3 out of compound means horse; but in historical fact, ma^3 is a recent borrowing (the first syllable of English 'macadamized'). When not followed by place expression, cai^4 appears in such euphemistic sentences as He's still living ([t'a^1] [cai^4]), as in answer to a question (whether one's father is still living); compare He's not in ([t'a^1] [pu^0 cai^4]) in response to asking whether someone is in his office. In some languages this morpheme alone (or alternating with čen^4 cai^4 in 'acquired Mandarin') precedes verbs of action in the sense of in the process of, as He's running ([t'a^1] [cai^4 p'au^3]). Interrogatives, and affirmation (can run fast) or negation of ability (can't run fast) in Han Chinese agree largely on the selection of morphemes (nonverbal negative for not or cannot, and te^0 or cognate for can), but differ in the order and cooccurrence of these morphemes.

(21) This kind can be eaten: topic functioning as object of verb is more commonly encountered in Cantonese than in Mandarin or Wu. In the latter languages, [this kind] as topic (demonstrative before classifier without following noun) usually functions as subject.

(22) Ambiguity as to whether topic functions as object or subject is entirely removed when empty morpheme for pause (ia^0) follows the topic as object: As for chicken (pause), (zero but potential embedded subject) is not eating any more. In relative order of subject (S), object (O), and verb (V), permutations that are possible in Han languages include SVO and OSV, SV and OV. The OV order is exemplified in As for chicken (O) not eating any more (with the subject not specified, but with possible ambiguity, since chicken might be taken as the sub-

ject). Stated positively, S may precede V (without specification of O); O may precede V (without specification of S); OS may precede V (in that order); O may follow V (with or without specification of S). Stated negatively S never follows V; S never follows V_1 in sentences with two link verbs, V_1 and V_2); O never precedes V_1 (in sentences with link verbs where S of V_1 precedes V_1 whose object follows; but this object functions as the subject of V_2 in pivotal O/S sequence S V_1 ←O/S→ V_2); OS before V is never inverted to SO before V (experimental attempts to say SO before V are understood as OS, since O and S are not marked for their function as object or subject except by their relative order in respect to V).

(23) The guest is just apt to come (can possibly come). The comment begins with huei^4 can possibly (which would be followed by the cognate of te^0 in Wu) which occurs, as auxiliary verb, with and without phrase nucleus verb, in such questions as Can you? [can] [not can] (huei^4 pu^0 huei^4) and Can you do it? [can do] [not can do] (huei^4 cuo^0 pu^1 huei^4 cuo^0). The topic functions as subject; guest is the translation unit of the compound (guest person). In Mandarin, the first member of the compound is free, while for (host person) the first member is bound — i.e. appears only in compound. But in Cantonese the cognates for guest as well as host are bound — found only in compounds.

(24) He has eaten noodles (as you can see). The perfective suffix after the verb (le^0) has the same shape as the sentence final suffix (le^0, out of bracket) in Mandarin; when both are used, as here, both meanings are marked, but ambiguity is possible when only one le^0 appears at the end of a sentence (as in 25). Such ambiguity is possible neither in Wu nor in Cantonese, since the two morphemes parallel in function to le^0 are entirely different in shape. And in general there is more differentiation of the kind suggested here in the

relatively small area of China in which Wu is spoken than in the three-fourths of China in which Mandarin is spoken — in particular, there is less differentiation in person pronouns in Mandarin dialects.

(25) With le^0 at the end of the sentence: if perfective, the sense is *The guest has come* (for le^0_1); if modal or situational sentence suffix, out of bracket, the sense is (*Oh look*), *a guest is coming* (for le^0_2). In Wu and Cantonese, the perfective suffix has one shape, the sentence situational suffix another, and the functions of the two are thereby kept distinct when either occurs at the end of a sentence.

(26) There appears to be no parallel in Cantonese to cai^4 *future again*, which occurs only with link verbs (28) or with final situational suffix; this usage is more common in Wu than in Mandarin, *He won't ever come again*, (*that's the situation*). When the order of the morphemes in the comment is changed from those marking *future again* *also* *not* to *also* *not* *future again*, a new sentence is formed: *He also is not coming again* (*that's it*).

(27) In Cantonese [*chat one stroke*] appears in place of the reduplication favored by Wu and SW Mandarin dialects for such sentences as *Let's chat a little more*.

(28) *I don't think he'll come again*. This sentence shows link verbs (for *think* and *come*); the linkage is effected by a bilateral pivot (a noun or, as here, a pronoun which serves as the object of the first link verb and as the subject of the second). Sentence profiles showing link verbs with bilateral pivots are found in all Han Chinese languages.

(29) *Ask him to come again* shows the bilateral pivot, $t'a^1$, functioning as object of *ask* and subject of *come*. Strings of such pivots may appear in one sentence, as that for *I ask you to ask him to come*.

(30) Topic as clause subject (utterable as self-sufficient sentence) before comment is universal for Han Chinese.

(31) *Are you going?* shows an [X] [*not* X] interrogative profile ([*go*] [*not go*]) in Mandarin. In Cantonese, when the verb morpheme or translation unit is bisyllabic, the first syllable precedes the morpheme for *not*, and both follow in an X *not* XY interrogative profile.

(32) *Do you want tea or coffee?* shows verb (V) and object (O) sequence before and after the phrase bracket for *or*: -[hai^2 šɨ^0] -, replaceable by - [šɨ^0] -, so that the interrogative profile is [VO$_1$] - [*or*] - [VO$_2$], which also appears in Wu and Cantonese. More briefly, in Mandarin, [VO$_1$] [VO$_2$] is an interrogative profile which omits the phrase bracket for [*or*] ; a bilaterally connecting phrase is never omitted in Wu or in Cantonese which distinguish morphemes for *either* from morphemes for *or*.

The interrogative profile [V$_2$O$_1$] - [*or*] - [V$_2$ O$_2$] appears after a transitive [V$_1$] in sentence (33): *Do you like to eat rice or eat noodles?* The transitive verb for *like* is not a co-verb with that for *to eat* because the subject of *like* is expressed in the topic (*you*), while its object is expressed in the interrogative profile, as given above. The Mandarin translation unit for *like* is a compound (*joy delight*); the cognate morphemes for this compound may be inverted in Wu (*delight joy*), as the translation unit for *to be glad*.

(34) *Is he jumping down?* The verb complement phrase -[*down away*] may be uttered once though at the end of the sentence, though this phrase is equally relevant to both alternatives in the interrogative profile [V] [*not* V] — at least so in Cantonese, Wu, and 'acquired Mandarin'. But in Peking Mandarin, the verb complements are incorporated in each alternative of the interrogative profile: [V *down away*] [*not* V *down away*].

Noun and pronoun (interrogative or personal) function in Han Chinese in general as subject before the verb and as object after the verb, as in (35) and (36).

(37) *Has he jumped down?* shows interrogative by final sentence suffix after the phrase bracket [*jump down away perfective*] , an order shared by Mandarin and Cantonese, but not by Wu. (so also, for sentence 40). It is possible in 'acquired Mandarin' to have the perfective verb [*jump perfective*] precede the verb complements [*down away*] ; this is the obligatory order in Wu.

(38) *Has he jumped down?* shows the interrogative by a two morpheme verbal negative phrase, [*not have*] , after [*jump down away perfective*]. In Cantonese, this interrogative profile appears with a one morpheme verbal negative; when this negative phrase is followed by an interrogative particle, out of the bracket, the interrogative is redundantly marked.

(39) The Mandarin etymon for *run* differs from that for Cantonese. The latter is cognate with that in Wenyen for *run* which also appears in Mandarin but then shifts in meaning to *walk*.

(40) See (37).

(41) See (37), for Wu and even Cantonese preference for having the equivalent of the perfective le^{0} after the verb, and before the verb complements (instead of the Mandarin order of [*eat down away perfective*] , or with quantified object — as for *an apple*, or *a bowl of rice* — after the perfective verb. The preferred Wu-Cantonese order (perfective after verb) for *he ate* and the preferred Mandarin order (perfective after verb complement or object) may both be used for *he has eaten.*

(42) *He arose* (in the sense of *getting up from bed* in Mandarin and Wu, but of *raising body* in Cantonese).

(43) and (44) show (č'i^3 lai^3) as a possible translation unit for *up* (directional complement), with verbs of action as *stand* (*He stood up*), *fly* (*He has flown up*). But the same morphemes in phrases with non-action or stationary verbs, as *laugh, think, talk*, mark another translation unit, *inceptive* (resultative complement), as in *He began to think, began to talk*, or (44) *He has started to laugh*. Either translation unit, the *directional* or *inceptive-resultative* is possible when this verb complement appears in phrases after verbs of action: *He has started to fly* (beside *flown up*). Hence, ambiguity is possible with verbs of action but not with non-active stationary verbs. Equivalence of two sequences or translation units for *up* and *begin, start* is widespread among Han Chinese languages.

Compare sentence (49), with the order V-*inceptive* O-*resultative* for *sing* (V) *song* (O) — *started to sing*; for the object (O) the title of a song may be given. This interlocking order is possible in all dialects, but is not equally favored. Compare (50) *He walked out (toward speaker)*, (51) *Put the money out* (with *money* as topic), (52) *He put the money out (toward speaker)*, with (53) *He took out a book*, in which the dual directional verb complement is interrupted by the object: [. . . *out*] [*one classifier book*]-[*hitherward*] . In Wu the dual verb complement (*out hitherward*) is repeated after the object [*one classifier book*] .

(55) *As for the bowl* (pause), *it got broken*. In Cantonese and Wu, instead of the morpheme for *broken* different morphemes with different etyma are selected (for *soggy* or *mushy* in Cantonese, and *pulverized* in Wu).

The Following Abbreviations Will Be Used

AA	American Anthropologist
ACLS	American Council of Learned Societies
AES-P	American Ethnological Society, Publication
AL	Anthropological Linguistics
APS-P	American Philosophical Society, Proceedings
APS-T	American Philosophical Society, Transactions
BAE-B	Bureau of American Ethnology, Bulletin
BAE-R	Bureau of American Ethnology, Report
CU	Columbia University Contributions to Anthropology
IJAL	International Journal of American Linguistics
IUPAL	Indiana University Publications in Anthropology and Linguistics
JAF	Journal of American Folklore
JSAP	Journal de la Société des Américanistes de Paris
Lg	Language
RCPAFL	Research Center Publications in Anthropology, Folklore and Linguistics
SJA	Southwestern Journal of Anthropology
SIL	Studies in Linguistics
TCLP	Travaux du Cercle Linguistique de Prague
UMPL	University of Michigan Publications, Linguistics
UCPAAE	University of California Publications in American Archaeology and Ethnology
UCPL	University of California Publications in Linguistics
VFPA	Viking Fund Publications in Anthropology
WDWLS	William Dwight Whitney Linguistic Series

CONTENTS

SINO-TIBETAN FASCICLE ONE

0.1. Authorship and sponsorship 1

0.2. Scope of Sino-Tibetan. 7

0.3. Previous Sino-Tibetan classifications 9

1.0. Han chinese as an ethnolinguistic monolith 14

1.1. Character writing. 19

1.2. Han Chinese in conversation and in Wenyen 29

1.3. Common national language = 'acquired Mandarin'. 34

1.4. Sentence Profiles of 'acquired Mandarin' 39

1.5. Diversity among Chinese languages and dialects 98